

SCALING LIMITS OF RANDOM GRAPHS FROM SUBCRITICAL CLASSES

KONSTANTINOS PANAGIOTOU, BENEDIKT STUFLER, AND KERSTIN WELLER

ABSTRACT. We study the uniform random graph C_n with n vertices drawn from a subcritical class of connected graphs. Our main result is that the rescaled graph C_n/\sqrt{n} converges to the Brownian Continuum Random Tree \mathcal{T}_e multiplied by a constant scaling factor that depends on the class under consideration. In addition, we provide subgaussian tail bounds for the diameter $D(C_n)$ and height $H(C_n^\bullet)$ of the rooted random graph C_n^\bullet . We give analytic expressions for the scaling factor of several classes, including for example the prominent class of outerplanar graphs. Our methods also enable us to study first passage percolation on C_n , where we show the convergence to \mathcal{T}_e under an appropriate rescaling.

1. INTRODUCTION

Let G be a connected graph with vertex set $V(G)$ and edge set $E(G)$. We can associate in a natural way a metric space $(V(G), d_G)$ to G , where $d_G(u, v)$ is the number of edges on a shortest path that contains u and v in G . In this work we study the case where G is a random graph, and we consider several properties of the associated metric space as the number of vertices of G becomes large.

In the series of seminal papers [3, 4, 5] Aldous studied the fundamental case of G being a critical Galton-Watson random tree with n vertices, where the offspring distribution has finite nonzero variance. Among other results, he showed that the asymptotic properties of the associated metric space admit an *universal* description: they can be depicted, up to an appropriate rescaling, in terms of “continuous trees” whose archetype is the so-called *Brownian Continuum Random Tree* (CRT for short). Since Aldous’s pioneering work, the CRT has been identified as the limiting object of many different classes of discrete structures, in particular trees, see e.g. Haas and Miermont [25] and references therein, and planar maps, see e.g. Albenque and Marckert [2], Bettinelli [9], Caraceni [12], Curien, Haas and Kortchemski [14] and Janson and Stefansson [28].

Although the aforementioned papers identify the CRT as the universal limiting object in various settings, much less is known about the scaling limit of random graphs from complex graph classes. In this paper we study in a unified way the asymptotic distribution of distances in random graphs from so-called *subcritical classes*, where, informally, a class is called subcritical if for a typical graph with n vertices the largest block (i.e. inclusion maximal 2-connected subgraph) has $O(\log n)$ vertices. Random graphs from such classes have been the object of intense research in the last years, see e.g. [18, 7, 17, 39], especially because of their close connection to the class of planar graphs. Prominent examples of classes that are subcritical are outerplanar and series-parallel graphs. However, with the notable exception of [18], most research on such random graphs has focused on *additive* parameters, like the

number of vertices of a given degree; the fine study of *global* properties, like the distribution of the distances, poses a significant challenge.

In the present paper we study the random graph C_n drawn uniformly from the set of connected graphs with n vertices of a subcritical class \mathcal{C} . Our first main result is Theorem 5.1, which shows that, up to an appropriate rescaling, the associated metric space converges in distribution to a multiple of the CRT. Postponing the introduction of the appropriate notation to later sections (see the outline), our main result asserts that there is a constant $s = s(\mathcal{C}) > 0$ such that

$$(V(C_n), s n^{-1/2} d_{C_n}) \xrightarrow{(d)} \mathcal{T}_e,$$

where \mathcal{T}_e is the CRT and convergence is with respect to the Gromov-Hausdorff metric. In particular, this establishes that the CRT is the universal scaling limit for random graphs from subcritical classes, and it proves (in a strong form) a conjecture by Drmota and Noy [18]. The proof of Theorem 5.1, see Section 5, gives a natural combinatorial characterization of the “scaling” constant s . Our methods are based on the algebraic concept of \mathcal{R} -enriched trees; more specifically, we use a size-biased enriched tree in order to study a coupling of C_n with an appropriate conditioned critical Galton-Watson tree T_n sharing the same vertex set. Our main step in the proof establishes that with high probability for any two vertices u, v in C_n the distance $d_{C_n}(u, v)$ is concentrated around $d_{T_n}(u, v)$ multiplied by a constant factor $\kappa \geq 1$ that depends *only* on \mathcal{C} , and which is, very roughly speaking, the average length of a shortest path between two random distinct vertices in a random block of C_n . Thus, the constant s turns out to be the product of two quantities: the constant involved in the scaling limit of T_n , and the reciprocal of κ . In Section 8 we exploit this characterization of s and compute its value for several important classes, including outerplanar graphs.

As a consequence of our main result we obtain the following statements, see Corollary 5.2. The *diameter* $D(G)$ of a graph G is defined as $\max_{u,v \in V(G)} d_G(u, v)$, and the *height* $H(G^\bullet)$ of a pointed graph $G^\bullet = (G, r)$, which is G rooted at a vertex r , is $\max_{v \in V(G)} d_G(r, v)$. Then the limit distribution for the diameter of C_n and the height of C_n^\bullet satisfy for $x > 0$, as $n \rightarrow \infty$

$$\begin{aligned} \mathbb{P}\left(D(C_n) > s^{-1} n^{1/2} x\right) &\rightarrow \sum_{k=1}^{\infty} (k^2 - 1) \left(\frac{2}{3} k^4 x^4 - 4k^2 x^2 + 2\right) \exp(-k^2 x^2/2), \\ \mathbb{P}\left(H(C_n^\bullet) > s^{-1} n^{1/2} x\right) &\rightarrow 2 \sum_{k=1}^{\infty} (4k^2 x^2 - 1) \exp(-2k^2 x^2). \end{aligned}$$

Apart from the convergence in distribution, we also show sharp tail bounds for the diameter and the height, see Theorem 6.1. In particular, we show that there are constants $C, c > 0$ such that *for all* n and $x \geq 0$

$$\mathbb{P}(D(C_n) \geq x) \leq C \exp(-cx^2/n) \quad \text{and} \quad \mathbb{P}(H(C_n^\bullet) \geq x) \leq C \exp(-cx^2/n).$$

A similar result was shown for critical Galton-Watson random trees by Addario-Berry, Devroye and Janson [1], and our proof of these bounds builds on the methods in that paper. From this we deduce that all moments of the rescaled height and diameter converge as well. In particular, we obtain the universal and remarkable asymptotic behaviour

$$\mathbb{E}[D(C_n)] \sim \frac{2^{3/2}}{3s} \sqrt{\pi n} \sim \frac{4}{3} \mathbb{E}[H(C_n^\bullet)].$$

This improves the previously best known bounds $c_1 \sqrt{n} \leq \mathbb{E}[D(C_n)] \leq c_2 \sqrt{n \log n}$ given in [18]. The higher moments can also be determined and are depicted in Sections 5 and 2.4.

In addition to the previous results, we demonstrate that our proof strategy is powerful enough to enable us to study the far more general setting of *first passage percolation*: suppose that the edges of \mathbb{C}_n are equipped with independent random “lengths”, drawn from a distribution that has exponential moments, and let the distance of two vertices u, v be the minimum sum of those lengths along a path that contains both u and v . Our last main result shows that again, up to an appropriate rescaling, the associated metric space converges to a multiple of the CRT; see Section 7 for the details.

Outline. Section 2 fixes some basic notation and summarizes several results related to Galton-Watson random trees and the Continuum Random Tree (CRT). In particular, Subsection 2.4 states the distribution and expressions for arbitrarily high moments of the height and diameter of the CRT – to our knowledge these results are scattered accross several papers and we provide a concise presentation. Section 3 is devoted to the definition of combinatorial species, \mathcal{R} -enriched trees and subcritical graph classes. In this part of the paper we collect some general and relevant properties of these objects – many of them were already known in special cases, and we put them in a broader context. In Section 4 we describe a construction of a novel powerful object called the *size-biased random enriched tree* that will allow us to study systematically the distribution of distances in random graphs from subcritical graph classes. Subsequently, in Section 5 we show our main result: the convergence of the rescaled random graphs towards a multiple of the CRT. Section 6 complements this result by proving subgaussian tail-bounds for the height and diameter. Section 7 is devoted to several extensions of our results, in particular first passage percolation. The paper closes with many examples, including the prominent class of outerplanar graphs.

2. GALTON-WATSON TREES AND THE CRT

We briefly summarize required notions and results related to the Brownian Continuum Random Tree (CRT) and refer the reader to Aldous [5] and Le Gall [34] for a thorough treatment.

2.1. Graphs and (Plane) Trees. All graphs considered in this paper are undirected and may not contain multiple edges or loops. That is, a graph G consists of a non-empty set $V(G)$ of vertices and a set $E(G)$ of edges that are two-element subsets of $V(G)$. If $|V(G)| \in \mathbb{N}$ we say that $|G| := |V(G)|$ is the *size* of G . Following Diestel [16], we recall and fix basic definitions and notations. Two vertices $v, w \in V(G)$ are said to be *adjacent* if $\{v, w\} \in E(G)$. We will often write vw instead of $\{v, w\}$. A *path* P is a graph such that

$$V(P) = \{v_0, \dots, v_\ell\}, \quad E(P) = \{v_0v_1, \dots, v_{\ell-1}v_\ell\}$$

with the v_i being distinct. The number $\ell = \ell(P)$ of edges of a path is its *length*. We say P *connects* its endvertices v_0 and v_ℓ and we often write $P = v_0v_1 \dots v_\ell$. If P has length at least two we call the graph $C_\ell = P + v_0v_\ell$ obtained by adding the edge v_0v_ℓ a *cycle*. The *complete graph* with n vertices in which each pair of distinct vertices is adjacent is denoted by K_n .

We say the graph G is *connected* if any two vertices $u, v \in V(G)$ are connected by a path in G . The length of a shortest path connecting the vertices u and v is called the *distance* of u and v and it is denoted by $d_G(u, v)$. Clearly d_G is a metric on the vertex set $V(G)$. A graph G together with a distinguished vertex $v \in V(G)$ is called a *rooted graph* with root-vertex v . The *height* $h(w)$ of a vertex $w \in V(G)$ is its distance from the root. The *height* $H(G)$ is the maximum height of the vertices in G . A *tree* T is a non-empty connected graph without cycles. Any two vertices of a tree are connected by a unique path. If T is rooted, then the

vertices $w' \in V(T)$ that are adjacent to a vertex w and have height $h(w) + 1$ form the *offspring* set of the vertex w . Its cardinality is the *outdegree* $d^+(w)$ of w .

The *Ulam-Harris tree* is an infinite rooted tree with vertex set $\cup_{n \geq 0} \mathbb{N}^n$ consisting of finite sequences of natural numbers. The empty string \emptyset is the root and the offspring of any vertex v is given by the concatenations $v1, v2, v3, \dots$. In particular, the labelling of the vertices induces a linear order on each offspring set. A *plane tree* is defined as a subtree of the Ulam-Harris tree that contains the root.

2.2. Galton-Watson Trees. Throughout this section we fix an integer-valued random variable $\xi \geq 0$. By abuse of language we will often not distinguish between ξ and its distribution. A ξ -*Galton-Watson tree* T is the family tree of a Galton-Watson branching process with offspring distribution ξ , interpreted as a (possibly infinite) plane tree. If $\mathbb{P}(\xi = 1) < 1$, then T is almost surely finite if and only if $\mathbb{E}[\xi] \leq 1$. If $\mathbb{E}[\xi] = 1$ we call T *critical*. Let $\text{supp}(\xi) = \{k \mid \mathbb{P}(\xi = k) > 0\}$ denote the *support* of ξ and define the *span* $d = \text{span}(\xi)$ as the greatest common divisor of $\{k - \ell \mid k, \ell \in \text{supp}(\xi)\}$. If T is finite, then

$$|\mathsf{T}| = 1 + \sum_{v \in V(\mathsf{T})} d_{\mathsf{T}}^+(v) \equiv 1 \pmod{d}.$$

Conversely, if ξ is not almost surely positive, then $\mathbb{P}(|\mathsf{T}| = n) > 0$ for all large enough $n \in \mathbb{N}$ satisfying $n \equiv 1 \pmod{d}$. We will need the following standard result for the probability that a critical Galton-Watson tree has size n .

Lemma 2.1 ([30, p. 105]). *Suppose that the distribution ξ has expected value one and finite nonzero variance $0 < \sigma^2 < \infty$. Let $(\xi_i)_{i \in \mathbb{N}}$ be an infinite family of i.i.d. copies of ξ . Then the probability that the ξ -Galton-Watson tree T has size $n \equiv 1 \pmod{d}$ with $d = \text{span}(\xi)$ satisfies*

$$\mathbb{P}(|\mathsf{T}| = n) = n^{-1} \mathbb{P}\left(\sum_{i=1}^n \xi_i = n - 1\right) \sim \frac{d}{\sqrt{2\pi\sigma^2}} n^{-3/2} \quad \text{as } n \rightarrow \infty.$$

We also state some results given in [1, 27] that will be useful in our arguments. As in the previous lemma, suppose that ξ satisfies $\mathbb{E}[\xi] = 1$ and has finite nonzero variance. Let $n \equiv 1 \pmod{d}$ be sufficiently large such that $\mathbb{P}(\mathsf{T} = n) > 0$ and let T_n denote the Galton-Watson tree conditioned on having size n . By [1, Thm. 1.2] and [1, p. 6] there are constants $C_1, c_1, C_2, c_2 > 0$ such that the height $H(\mathsf{T}_n)$ satisfies the following tail bounds for all n and $h \geq 0$:

$$(2.1) \quad \mathbb{P}(H(\mathsf{T}_n) \leq h) \leq C_1 \exp(-c_1(n-2)/h^2), \quad \mathbb{P}(H(\mathsf{T}_n) \geq h) \leq C_2 \exp(-c_2 h^2/n).$$

2.3. Convergence of Rescaled Conditioned Galton-Watson trees. *Throughout this and the following subsections in Section 2 we write T for a critical ξ -Galton-Watson tree having finite nonzero variance σ^2 . Moreover, n will always denote an integer satisfying $n \equiv 1 \pmod{\text{span}(\xi)}$ and is assumed to be large enough such that the conditioned tree T_n having exactly n vertices is well-defined.*

Given a plane tree T of size n we consider its canonical *depth-first search* walk $(v_i)_{0 \leq i \leq 2(n-1)}$ that starts at the root and always traverses the leftmost unused edge first. That is, v_0 is the root of T and given v_0, \dots, v_i walk if possible to the leftmost unvisited son of v_i . If v_i has no sons or all sons have already been visited, then try to walk to the parent of v_i . If this is not possible either, being only the case when v_i is the root of T and all other vertices have already been visited, then terminate the walk. The corresponding heights $c(i) := d(\text{root of } T, v_i)$

define the search-depth function c of the tree T . The *contour function* $C : [0, 2(n-1)] \rightarrow \mathbb{R}_+$ is defined by $C(i) = c(i)$ for all integers $0 \leq i \leq 2(n-1)$ with linear interpolation between these values, see Figure 1 for an example. It can be shown that after a suitable rescaling the

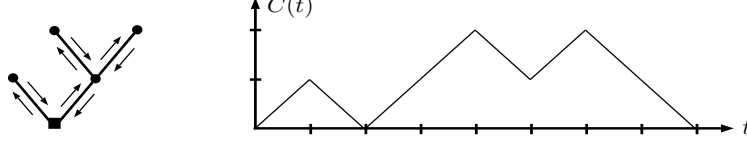


FIGURE 1. The contour function of a plane tree.

contour process of T_n converges to a normalized Brownian excursion.

Theorem 2.2 ([35, Thm 6.1]). *Let T_n be a critical ξ -Galton-Watson tree conditional on having n vertices, where ξ has finite non-zero variance σ^2 . Let C_n denote the contour function of T_n . Then*

$$\left(\frac{\sigma}{2\sqrt{n}} C_n(t2(n-1)) \right)_{0 \leq t \leq 1} \xrightarrow{(d)} \mathbf{e}$$

in $\mathcal{C}([0, 1], \mathbb{R}_+)$, where $\mathbf{e} = (\mathbf{e}_t)_{0 \leq t \leq 1}$ denotes the Brownian excursion of duration one.

This result is due to Aldous [5, Thm. 23], who stated it for aperiodic offspring distributions. See also Duquesne [20] and Kortchemski [31] for further extensions. Theorem 2.2 can be formulated as a convergence of random trees with respect to the Gromov-Hausdorff metric. We introduce the required notions following Le Gall and Miermont [36].

A continuous function $g : [0, 1] \rightarrow [0, \infty)$ with $g(0) = g(1) = 0$ induces a pseudo-metric on the interval $[0, 1]$ by

$$d_g(u, v) = g(u) + g(v) - 2 \inf_{u \leq s \leq v} g(s)$$

for $u \leq v$. This defines a metric on the quotient $T_g = [0, 1]/\sim$, where $u \sim v$ if and only if $d(u, v) = 0$; we denote the corresponding metric space by (T_g, d_g) and call the equivalence class $r_0(T_g)$ of the origin its *root*.

Given a metric space (E, d) the set of compact subsets is again a metric space with respect to the *Hausdorff metric*

$$\delta_H(K, K') = \inf\{\epsilon > 0 \mid K \subset U_\epsilon(K'), K' \subset U_\epsilon(K)\},$$

where $U_\epsilon(K) = \{x \in E \mid d(x, K) \leq \epsilon\}$. The sets \mathbb{K} (\mathbb{K}^\bullet) of isometry classes of (pointed) compact metric spaces, where a pointed space is a triple (E, d, r) where (E, d) is a metric space and $r \in E$ is a distinguished element, are Polish spaces with respect to the (*pointed*) *Gromov-Hausdorff metric*

$$d_{\text{GH}}((E_1, d_1), (E_2, d_2)) = \inf_{\varphi_1, \varphi_2} \{\epsilon > 0 \mid \delta_H(\varphi_1(E_1), \varphi_2(E_2)) \leq \epsilon\},$$

$$d_{\text{GH}}((E_1, d_1, r_1), (E_2, d_2, r_2)) = \inf_{\varphi_1, \varphi_2} \left\{ \epsilon > 0 \mid \max \left\{ \delta_H(\varphi_1(E_1), \varphi_2(E_2)), d_E(\varphi_1(r_1), \varphi_2(r_2)) \right\} \leq \epsilon \right\},$$

where the infimum is in both cases taken over all isometric embeddings $\varphi_1 : E_1 \rightarrow E$ and $\varphi_2 : E_2 \rightarrow E$ into a common metric space (E, d_E) [36, Thm. 3.5]. We will make use of the following characterisation of the Gromov-Hausdorff metric. Given two compact metric spaces (E_1, d_1) and (E_2, d_2) a *correspondence* between them is a subset $R \subset E_1 \times E_2$ such that any point $x \in E_1$ corresponds to at least one point $y \in E_2$ and vice versa. The *distortion* of R is

$$\text{dis}(R) = \sup\{|d_1(x_1, y_1) - d_2(x_2, y_2)| \mid (x_1, y_1), (x_2, y_2) \in R\}.$$

Proposition 2.3 ([36, Prop. 3.6]). *Given two pointed compact metric spaces (E_1, d_1, r_1) and (E_2, d_2, r_2) we have that*

$$2d_{GH}((E_1, d_1, r_1), (E_2, d_2, r_2)) = \inf_R \text{dis}(R),$$

where R ranges over all correspondences between E_1 and E_2 such that r_1, r_2 correspond to each other.

An important consequence is that the mapping

$$(\{g \in \mathcal{C}([0, 1], \mathbb{R}_+) \mid g(0) = g(1) = 0\}, \|\cdot\|_\infty) \rightarrow (\mathbb{K}, d_{GH}), \quad g \mapsto T_g$$

is Lipschitz-continuous [36, Cor. 3.7].

Definition 2.4. *The random metric space $(\mathcal{T}_e, d_e, r_0(\mathcal{T}_e))$ coded by the Brownian excursion of duration one e is called the Brownian continuum random tree (CRT).*

We may view T_n as a random pointed metric space $(V(T_n), d_{T_n}, \emptyset) \in \mathbb{K}$. This space is close to the real tree encoded by its contour function, hence Theorem 2.2 implies convergence with respect to the Gromov-Hausdorff metric, see [35, p. 740].

Theorem 2.5. *Let T_n be a critical ξ -Galton-Watson tree conditional on having n vertices, where ξ has finite non-zero variance σ^2 . As n tends to infinity, T_n with edges rescaled to length $\frac{\sigma}{2\sqrt{n}}$ converges in distribution to the CRT, that is*

$$(2.2) \quad \left(V(T_n), \frac{\sigma}{2\sqrt{n}} d_{T_n}, \emptyset \right) \xrightarrow{(d)} (\mathcal{T}_e, d_e, r_0(\mathcal{T}_e))$$

in the metric space $(\mathbb{K}^\bullet, d_{GH})$.

This invariance principle is due to Aldous [5] and there exist various extensions. See for example Duquesne [20], Duquesne and Le Gall [21], Haas and Miermont [25]. Adopting common terminology, we will often write in the sequel

$$\frac{\sigma}{2\sqrt{n}} T_n \xrightarrow{(d)} \mathcal{T}_e$$

instead of (2.2).

2.4. Height and Diameter of the CRT. Clearly the height $H(T_n)$ and Diameter $D(T_n)$ of the tree T_n may be recovered from its contour function C . By the results from Section 2.3 it follows that

$$(2.3) \quad \frac{\sigma}{2\sqrt{n}} H(T_n) \xrightarrow{(d)} H(\mathcal{T}_e) \stackrel{(d)}{=} \sup_{0 \leq t \leq 1} e(t)$$

$$(2.4) \quad \frac{\sigma}{2\sqrt{n}} D(T_n) \xrightarrow{(d)} D(\mathcal{T}_e) \stackrel{(d)}{=} \sup_{0 \leq t_1 \leq t_2 \leq 1} (e(t_1) + e(t_2) - 2 \inf_{t_1 \leq t \leq t_2} e(t)).$$

Since $D(T_n) \leq 2H(T_n)$ the tail bound (2.1) implies that all moments in (2.3) and (2.4) converge. It is well-known that $H(\mathcal{T}_e)/\sqrt{2}$ follows a *Theta distribution*, that is

$$(2.5) \quad \mathbb{P}(H(\mathcal{T}_e) > x) = 2 \sum_{k=1}^{\infty} (4k^2 x^2 - 1) \exp(-2k^2 x^2).$$

for all $x > 0$. The k th moment of the height is given by

$$(2.6) \quad \mathbb{E}[H(\mathcal{T}_e)] = \sqrt{\pi/2} \quad \text{and} \quad \mathbb{E}[H(\mathcal{T}_e)^k] = 2^{-k/2} k(k-1) \Gamma(k/2) \zeta(k) \quad \text{for } k \geq 2.$$

This follows from standard results on the Brownian excursion, see e.g. Chung [13] and Biane, Pitman and Yor [10], or by calculating directly the limit distribution of extremal parameters of a class of trees that converges to the CRT (see e.g. Rényi and Szekeres [40]). The distribution of the diameter is given by

$$(2.7) \quad \mathbb{P}(D(\mathcal{T}_e) > x) = \sum_{k=1}^{\infty} (k^2 - 1) \left(\frac{2}{3} k^4 x^4 - 4k^2 x^2 + 2 \right) \exp(-k^2 x^2 / 2).$$

This expression may be obtained (by tedious calculations) from results of Szekeres [42], who proved the existence of a limit distribution for the diameter of rescaled random unordered labelled trees. The moments of this distribution were calculated for example in Broutin and Flajolet [11] and are given by

$$(2.8) \quad \mathbb{E}[D(\mathcal{T}_e)] = \frac{4}{3} \sqrt{\pi/2}, \quad \mathbb{E}[D(\mathcal{T}_e)^2] = \frac{2}{3} \left(1 + \frac{\pi^2}{3} \right), \quad \mathbb{E}[D(\mathcal{T}_e)^3] = 2\sqrt{2\pi},$$

$$(2.9) \quad \mathbb{E}[D(\mathcal{T}_e)^k] = \frac{2^{k/2}}{3} k(k-1)(k-3) \Gamma(k/2) (\zeta(k-2) - \zeta(k)) \quad \text{for } k \geq 4.$$

3. COMBINATORIAL SPECIES, \mathcal{R} -ENRICHED TREES AND SUBCRITICAL GRAPH CLASSES

We recall parts of the theory of combinatorial species and Boltzmann samplers to the extent required in this paper. A reader who is already familiar with the framework of subcritical graph classes may skip some parts of this section. However, we stress the importance of the representation of connected graphs as enriched trees in Subsection 3.5 and the coupling of random graphs with a Galton-Watson tree in Subsection 3.7. Moreover, several intermediate lemmas that we state were already shown in previous papers, albeit under stronger assumptions.

3.1. Combinatorial Species. The framework of combinatorial species allows for a unified treatment of a wide range of combinatorial objects. We give only a concise introduction and refer to Joyal [29] and Bergeron, Labelle and Leroux [6] for a detailed discussion. The essentially equivalent language of *combinatorial classes* was developed by Flajolet and Sedgewick [23].

A *combinatorial species* may be defined as a functor \mathcal{F} that maps any finite set U of *labels* to a finite set $\mathcal{F}[U]$ of \mathcal{F} -objects and any bijection $\sigma : U \rightarrow V$ of finite sets to its (bijective) *transport function* $\mathcal{F}[\sigma] : \mathcal{F}[U] \rightarrow \mathcal{F}[V]$ along σ , such that composition of maps and the identity are preserved. We say that a species \mathcal{G} is a *subspecies* of \mathcal{F} and write $\mathcal{G} \subset \mathcal{F}$ if $\mathcal{G}[U] \subset \mathcal{F}[U]$ for all finite sets U and $\mathcal{G}[\sigma] = \mathcal{F}[\sigma]|_{\mathcal{G}[U]}$ for all bijections $\sigma : U \rightarrow V$. Given two species \mathcal{F} and \mathcal{G} , an *isomorphism* $\alpha : \mathcal{F} \xrightarrow{\sim} \mathcal{G}$ from \mathcal{F} to \mathcal{G} is a family of bijections $\alpha = (\alpha_U : \mathcal{F}[U] \rightarrow \mathcal{G}[U])_U$ where U ranges over all finite sets, such that $\mathcal{G}[\sigma]\alpha_U = \alpha_V \mathcal{F}[\sigma]$ for all bijective maps $\sigma : U \rightarrow V$. The species \mathcal{F} and \mathcal{G} are *isomorphic* if there exists an isomorphism from one to the other. This is denoted by $\mathcal{F} \simeq \mathcal{G}$ or, by abuse of notation, just $\mathcal{F} = \mathcal{G}$. An element $\gamma_U \in \mathcal{F}[U]$ has size $|\gamma_U| := |U|$ and two \mathcal{F} -objects γ_U and γ_V are termed isomorphic if there is a bijection $\sigma : U \rightarrow V$ such that $\mathcal{F}[\sigma](\gamma_U) = \gamma_V$. We will often just write $\sigma.\gamma_U = \gamma_V$ instead, if there is no risk of confusion. An isomorphism class of \mathcal{F} -structures is called an *unlabelled \mathcal{F} -object*.

We will mostly be interested in subspecies of the species of finite simple graphs and use basic species such as the species of linear orders SEQ or the SET -species defined by $\text{SET}[U] = \{U\}$ for all U . Moreover let 0 denote the empty species, 1 the species with a single object of size 0 and \mathcal{X} the species with a single object of size 1.

We set $[n] := \{1, \dots, n\}$ and $\mathcal{F}_n := \mathcal{F}[n] = \mathcal{F}[\{1, \dots, n\}]$ for all $n \in \mathbb{N}_0$. By abuse of notation we will often let \mathcal{F} also refer to the set $\cup_n \mathcal{F}_n$. The *exponential generating series* of a combinatorial species \mathcal{F} is defined by $F(x) = \sum_{n \geq 0} |\mathcal{F}_n| x^n / n!$. In general, this is a formal power series that may have radius of convergence zero. If the series $F(x)$ has positive radius of convergence, we say that \mathcal{F} is an *analytic species* and $F(x)$ is its *exponential generating function*. For any power series $f(x)$ we let $[x^n]f(x)$ denote the coefficient of x^n .

3.2. Operations on Species. The framework of combinatorial species offers a large variety of constructions that create new species from others. In the following let \mathcal{F} , $(\mathcal{F}_i)_{i \in \mathbb{N}}$ and \mathcal{G} denote species and U an arbitrary finite set. The *sum* $\mathcal{F} + \mathcal{G}$ is defined by the disjoint union

$$(\mathcal{F} + \mathcal{G})[U] = \mathcal{F}[U] \sqcup \mathcal{G}[U].$$

More generally, the infinite sum $(\sum_i \mathcal{F}_i)$ may be defined by $(\sum_i \mathcal{F}_i)[U] = \bigsqcup_i \mathcal{F}_i[U]$ if the right hand side is finite for all finite sets U . The *product* $\mathcal{F} \cdot \mathcal{G}$ is defined by the disjoint union

$$(\mathcal{F} \cdot \mathcal{G})[U] = \bigsqcup_{\substack{(U_1, U_2) \\ U_1 \cap U_2 = \emptyset, U_1 \cup U_2 = U}} \mathcal{F}[U_1] \times \mathcal{G}[U_2]$$

with componentwise transport. Thus, n -sized objects of the product are pairs of \mathcal{F} -objects and \mathcal{G} -objects whose sizes add up to n . If the species \mathcal{G} has no objects of size zero, we can form the *substitution* $\mathcal{F} \circ \mathcal{G}$ by

$$(\mathcal{F} \circ \mathcal{G})[U] = \bigsqcup_{\pi \text{ partition of } U} \mathcal{F}[\pi] \times \prod_{Q \in \pi} \mathcal{G}[Q].$$

An object of the substitution may be interpreted as an \mathcal{F} -object whose labels are substituted by \mathcal{G} -objects. The transport along a bijection σ is defined by applying the induced map $\bar{\sigma} : \pi \rightarrow \bar{\pi} = \{\sigma(Q) \mid Q \in \pi\}$ of partitions to the \mathcal{F} -object and the restricted maps $\sigma|_Q$ with $Q \in \pi$ to their corresponding \mathcal{G} -objects. We will often write $\mathcal{F}(\mathcal{G})$ instead of $\mathcal{F} \circ \mathcal{G}$. The *rooted* or *pointed* \mathcal{F} -species is given by

$$\mathcal{F}^\bullet[U] = U \times \mathcal{F}[U]$$

with componentwise transport. That is, a pointed object is formed by distinguishing a label, named the *root* of the object, and any transport function is required to preserve roots. The *derived* species \mathcal{F}' is defined by

$$\mathcal{F}'[U] = \mathcal{F}[U \cup \{*_U\}]$$

with $*_U$ referring to an arbitrary fixed element not contained in the set U . (For example, we could take $*_U = U$.) The transport along a bijective map $\sigma : U \rightarrow V$ is done by applying the canonically extended bijection $\sigma' : U \sqcup \{*_U\} \rightarrow V \sqcup \{*_V\}$ with $\sigma'(*_U) = *_V$ to the object. Derivation and pointing are related by an isomorphism $\mathcal{F}^\bullet \simeq \mathcal{F}'$. Note that \mathcal{F}^\bullet and \mathcal{F}' are in general different species, since a \mathcal{F}^\bullet -object may be rooted only at a non- $*$ -label.

Explicit formulas for the exponential generating series of these constructions are summarized in Table 1. The notation is quite suggestive: up to (canonical) isomorphism, each operation considered in this section is associative. The sum and product are commutative operations and satisfy a distributive law, i.e.

$$(3.1) \quad \mathcal{F} \cdot (\mathcal{G}_1 + \mathcal{G}_2) \simeq \mathcal{F} \cdot \mathcal{G}_1 + \mathcal{F} \cdot \mathcal{G}_2.$$

$\sum_i \mathcal{F}_i$	$\sum_i F_i(x)$	SET	$\exp(x)$
$\mathcal{F} \cdot \mathcal{G}$	$F(x)G(x)$	SEQ	$1/(1-x)$
$\mathcal{F} \circ \mathcal{G}$	$F(G(x))$	0	0
\mathcal{F}^\bullet	$x \frac{d}{dx} F(x)$	1	1
\mathcal{F}'	$\frac{d}{dx} F(x)$	\mathcal{X}	x

TABLE 1. Relation between combinatorial constructions and generating series.

The operation of deriving a species is additive and satisfies a product rule and a chain rule, analogous to the derivative in calculus:

$$(3.2) \quad (\mathcal{F} \cdot \mathcal{G})' \simeq \mathcal{F}' \cdot \mathcal{G} + \mathcal{F} \cdot \mathcal{G}' \quad \text{and} \quad \mathcal{F}(\mathcal{G})' \simeq \mathcal{F}'(\mathcal{G}) \cdot \mathcal{G}'.$$

Recall that for the chain rule to apply we have to require $\mathcal{G}[\emptyset] = \emptyset$, since otherwise $\mathcal{F}(\mathcal{G})$ is not defined. A thorough discussion of these facts is beyond the scope of this introduction. We refer the inclined reader to Joyal [29] and Bergeron, Labelle and Leroux [6].

3.3. Combinatorial Specifications. In this section we briefly recall Joyal’s implicit species theorem that allows us to define combinatorial species up to unique isomorphism and construct recursive samplers that draw objects of a species randomly (see Section 3.6 below). In order to state the theorem we need to introduce the concept of *multisort species*. As it is sufficient for our applications, we restrict ourselves to the 2-sort case.

A *2-sort species* \mathcal{H} is a functor that maps any pair $U = (U_1, U_2)$ of finite sets to a finite set $\mathcal{H}[U] = \mathcal{H}[U_1, U_2]$ and any pair $\sigma = (\sigma_1, \sigma_2)$ of bijections $\sigma_i : U_i \rightarrow V_i$ to a bijection $\mathcal{H}[\sigma] : \mathcal{H}[U] \rightarrow \mathcal{H}[V]$ in such a way, that identity maps and composition of maps are preserved. The operations of sum, product and composition extend naturally to the multisort-context. Let \mathcal{H} and \mathcal{K} be 2-sort species and $U = (U_1, U_2)$ a pair of finite sets. The *sum* is defined by

$$(\mathcal{H} + \mathcal{K})[U] = \mathcal{H}[U] \sqcup \mathcal{K}[U].$$

We write $U = V + W$ if $U_i = V_i \cup W_i$ and $V_i \cap W_i = \emptyset$ for all i . The *product* is defined by

$$(\mathcal{H} \cdot \mathcal{K})[U] = \bigsqcup_{V+W=U} \mathcal{H}[V] \times \mathcal{K}[W].$$

The *partial derivatives* are given by

$$\partial_1 \mathcal{H}[U] = \mathcal{H}[U_1 \cup \{*_U\}, U_2] \quad \text{and} \quad \partial_2 \mathcal{H}[U] = \mathcal{H}[U_1, U_2 \cup \{*_U\}].$$

In order state Joyal’s implicit species theorem we also require the substitution operation for multisort species; this will allow us to define species “recursively” up to (canonical) isomorphism. Let \mathcal{F}_1 and \mathcal{F}_2 be (1-sort) species and M a finite set. A structure of the *composition* $\mathcal{H}(\mathcal{F}_1, \mathcal{F}_2)$ over the set M is a quadrupel $(\pi, \chi, \alpha, \beta)$ such that:

- (1) π is partition of the set M .
- (2) $\chi : \pi \rightarrow \{1, 2\}$ is a function assigning to each class a sort.
- (3) α a function that assigns to each class $Q \in \pi$ a $\mathcal{F}_{\chi(Q)}$ object $\alpha(Q) \in \mathcal{F}_{\chi(Q)}[Q]$.
- (4) β a \mathcal{H} -structure over the pair $(\chi^{-1}(1), \chi^{-1}(2))$.

This construction is *functorial*: any pair of isomorphisms α_1, α_2 with $\alpha_i : \mathcal{F}_i \xrightarrow{\sim} \mathcal{G}_i$ induces an isomorphism $\mathcal{H}[\alpha_1, \alpha_2] : \mathcal{H}(\mathcal{F}_1, \mathcal{F}_2) \xrightarrow{\sim} \mathcal{H}(\mathcal{G}_1, \mathcal{G}_2)$.

Let \mathcal{H} be a 2-sort species and recall that \mathcal{X} denotes the species with a unique object of size one. A solution of the system $\mathcal{Y} = \mathcal{H}(\mathcal{X}, \mathcal{Y})$ is pair (\mathcal{A}, α) of a species \mathcal{A} with $\mathcal{A}[0] = 0$ and

an isomorphism $\alpha : \mathcal{A} \xrightarrow{\sim} \mathcal{H}(\mathcal{X}, \mathcal{A})$. An isomorphism of two solutions (\mathcal{A}, α) and (\mathcal{B}, β) is an isomorphism of species $u : \mathcal{A} \xrightarrow{\sim} \mathcal{B}$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\alpha} & \mathcal{H}(\mathcal{X}, \mathcal{A}) \\ \downarrow u & & \downarrow \mathcal{H}(\text{id}, u) \\ \mathcal{B} & \xrightarrow{\beta} & \mathcal{H}(\mathcal{X}, \mathcal{B}) \end{array}$$

We may now state Joyal's implicit species theorem.

Theorem 3.1 ([29], Théorème 6). *Let \mathcal{H} be a 2-sort species satisfying $\mathcal{H}(0, 0) = 0$. If $(\partial_2 \mathcal{H})(0, 0) = 0$, then the system $\mathcal{Y} = \mathcal{H}(\mathcal{X}, \mathcal{Y})$ has up to isomorphism only one solution. Moreover, between any two given solutions there is exactly one isomorphism.*

We say that an isomorphism $\mathcal{F} \simeq \mathcal{H}(\mathcal{X}, \mathcal{F})$ is a *combinatorial specification* for the species \mathcal{F} if the 2-sort species \mathcal{H} satisfies the requirements of Theorem 3.1.

3.4. Block-Stable Graph Classes. Any graph may be decomposed into its *connected components*, i.e. its maximal connected subgraphs. These connected components allow a *block-decomposition* which we recall in the following. Let C be a connected graph. If removing a vertex v (and deleting all adjacent edges) disconnects the graph, we say that v is a *cutvertex* of C . The graph C is *2-connected*, if it has size at least three and no cutvertices.

A *block* of an arbitrary graph G is a maximal connected subgraph $B \subset G$ that does not have a cutvertex (of itself). It is well-known, see for example [16], that any block is either 2-connected or an edge or a single isolated point. Moreover, the intersection of two blocks is either empty or a cutvertex of a connected component of G . If G is connected, then the bipartite graph whose vertices are the blocks and the cutvertices of G and whose edges are pairs $\{v, B\}$ with $v \in B$ is a tree and called the *block-tree* of G .

Let \mathcal{G} denote a subspecies of the species of graphs, $\mathcal{C} \subset \mathcal{G}$ the subspecies of connected graphs in \mathcal{G} and $\mathcal{B} \subset \mathcal{C}$ the subspecies of all graphs in \mathcal{C} , that are 2-connected or consist of only two vertices joined by an edge. We say that \mathcal{G} or \mathcal{C} is a *block-stable* class of graphs, if $\mathcal{B} \neq 0$ and $G \in \mathcal{G}$ if and only if every block of G belongs to \mathcal{B} or is a single isolated vertex. Block-stable classes satisfy the following combinatorial specifications that can be found for example in Joyal [29], Bergeron, Labelle and Leroux [6] and Harary and Palmer [26]:

$$(3.3) \quad \mathcal{G} \simeq \text{SET} \circ \mathcal{C} \quad \text{and} \quad \mathcal{C}^\bullet \simeq \mathcal{X} \cdot (\text{SET} \circ \mathcal{B}' \circ \mathcal{C}^\bullet).$$

The first correspondence expresses the fact that we may form any graph on a given vertex set U by partitioning U and constructing a connected graph on each partition class. The specification for rooted connected graphs, illustrated in Figure 2, is based on the construction of the block-tree. The idea is to interpret $\mathcal{B}' \circ \mathcal{C}^\bullet$ -objects as graphs by connecting the roots of the \mathcal{C}^\bullet objects on the partition classes and the $*$ -vertex with edges according to the \mathcal{B}' -object on the partition. An object of $\text{SET} \circ (\mathcal{B}' \circ \mathcal{C}^\bullet)$ can then be interpreted as a graph by identifying the $*$ -vertices of the $\mathcal{B}' \circ \mathcal{C}^\bullet$ objects. This construction is compatible with graph isomorphisms, hence $\mathcal{C}' \simeq \text{SET} \circ \mathcal{B}' \circ \mathcal{C}^\bullet$ and the second specification in (3.3) follows. By the rules for computing the generating series of species we obtain the equations

$$(3.4) \quad G(x) = \exp(C(x)) \quad \text{and} \quad C^\bullet(x) = x \exp(B'(C^\bullet(x))).$$

The following lemma was given in Panagiotou and Steger [39] and Drmota et al. [17] under some minor additional assumptions.

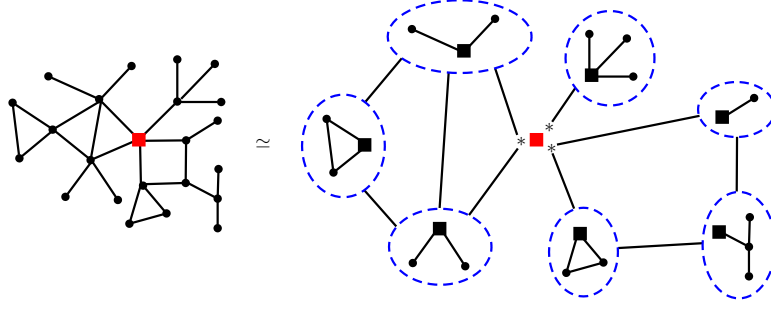


FIGURE 2. Decomposition of a rooted graph from \mathcal{C}^\bullet into a $\mathcal{X} \cdot (\text{SET} \circ \mathcal{B}' \circ \mathcal{C}^\bullet)$ structure. Labels are omitted and the roots are marked with squares.

Lemma 3.2. *Let \mathcal{C} be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. Then the exponential generating series $C(z)$ has radius of convergence $\rho < \infty$ and the sums $y := C^\bullet(\rho)$ and $\lambda := B'(y)$ are finite and satisfy*

$$(3.5) \quad y = \rho \exp(\lambda).$$

Proof. It suffices to consider the case $\rho > 0$. By assumption we have $\mathcal{B} \neq 0$ and hence there is a $k \in \mathbb{N}$ such that $[z^k]B'(z) \neq 0$. Thus, by (3.4) we have, say, $C^\bullet(z) = czC^\bullet(z)^{2k} + R(z)$ for some constant $c > 0$ and $R(z)$ a power series in z with nonnegative coefficients. This implies $\lim_{x \uparrow \rho} C^\bullet(x) < \infty$ and thus ρ and $C^\bullet(\rho)$ are both finite. The coefficients of all power series involved in (3.4) are nonnegative, and so it follows that $y = \rho \exp(\lambda)$ and thus $\lambda < \infty$. \square

We will only be interested in the case where \mathcal{C} is analytic. The following observation (made for example also in [18]) shows that this is equivalent to requiring that \mathcal{B} is analytic. We include a short proof for completeness.

Proposition 3.3. *Let \mathcal{C} be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. Then \mathcal{C} is analytic if and only if \mathcal{B} is analytic.*

Proof. By nonnegativity of coefficients we see easily that $\rho > 0$ implies that \mathcal{B} is analytic. Conversely, suppose that $B(z)$ has positive radius of convergence $R > 0$. By the inverse function theorem, the block-stability equation $f(z) = z \exp(B'(f(z)))$ has an analytic solution whose expansion at the point 0 agrees with the series $C^\bullet(z)$ by Lagrange's inversion formula. Hence \mathcal{C} is an analytic class. \square

3.5. \mathcal{R} -enriched Trees. The class \mathcal{A} of rooted trees¹ is known to satisfy the decomposition

$$\mathcal{A} \simeq \mathcal{X} \cdot \text{SET}(\mathcal{A}).$$

This is easy to see: in order to form a rooted tree on a given set of vertices, we choose a root vertex v , partition the remaining vertices, endow each partition class with a structure of a rooted tree and connect the vertex v with their roots. More generally, given a species \mathcal{R} the class $\mathcal{A}_{\mathcal{R}}$ of \mathcal{R} -enriched trees is defined by the combinatorial specification

$$\mathcal{A}_{\mathcal{R}} \simeq \mathcal{X} \cdot \mathcal{R}(\mathcal{A}_{\mathcal{R}}).$$

¹ *Arborescence* is the French word for rooted tree, hence the notation \mathcal{A} .

In other words, a \mathcal{R} -enriched tree is a rooted tree such that the offspring set of any vertex is endowed with a \mathcal{R} -structure. Natural examples are labeled ordered trees, which are SEQ -enriched trees, and plane trees, which are unlabeled ordered trees. Ordered and unordered tree families defined by restrictions on the allowed outdegree of internal vertices also fit in this framework. \mathcal{R} -enriched trees were introduced by Labelle [32] in order to provide a combinatorial proof of Lagrange Inversion. They have applications in various fields of mathematics, see for example [38, 15, 33].

The combinatorial specification (3.3) together with Theorem 3.1 allows us to identify a block-stable graph class \mathcal{C}^\bullet with the class \mathcal{R} -enriched trees where $\mathcal{R} = \text{SET}(\mathcal{B}')$, that is, rooted trees from \mathcal{A} where the offspring set of each vertex is partitioned into nonempty sets and each of these sets carries a \mathcal{B}' -structure. Compare with Figure 3.

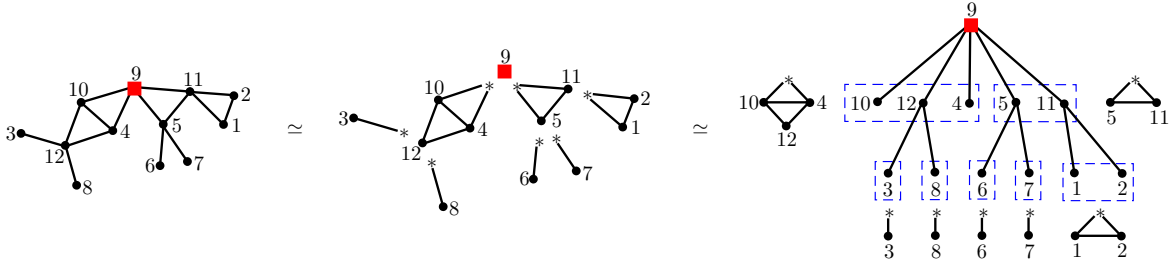


FIGURE 3. Correspondence of the classes \mathcal{C}^\bullet and $\text{SET}(\mathcal{B}')$ -enriched trees.

Corollary 3.4. *Let \mathcal{C} be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. Then there is a unique isomorphism between \mathcal{C}^\bullet and the class $\mathcal{A}_{\text{SET} \circ \mathcal{B}'}$ of pairs (T, α) with $T \in \mathcal{A}$ and α a function that assigns to each $v \in V(T)$ a (possibly empty) set $\alpha(v) \in (\text{SET} \circ \mathcal{B}')[M_v]$ of derived blocks whose vertex sets partition the offspring set M_v of v .*

Proof. By the isomorphism given in (3.3) the classes $\mathcal{A}_{\text{SET} \circ \mathcal{B}'}$ and \mathcal{C}^\bullet are both solutions of the system $\mathcal{Y} = \mathcal{H}(\mathcal{X}, \mathcal{Y})$ with $\mathcal{H}(\mathcal{X}, \mathcal{Y}) = \mathcal{X} \cdot \text{SET} \circ \mathcal{B}' \circ \mathcal{Y}$. Joyal's Implicit Species Theorem 3.1 yields that there is a unique isomorphism between any two solutions. \square

3.6. Boltzmann Samplers. Boltzmann samplers provide a method of generating efficiently random discrete combinatorial objects. They were introduced in Duchon, Flajolet, Louchard, and Schaeffer [19] and were developed further in Flajolet, Fusy and Pivoteau [22]. Following these sources we will briefly recall the theory of Boltzmann samplers to the extent required for the applications in this paper. Let $\mathcal{F} \neq 0$ be an analytic species of structures and F its exponential generating function. Given a parameter $x > 0$ such that $0 < F(x) < \infty$, a Boltzmann sampler $\Gamma F(x)$ is a random generator that draws an object $\gamma \in \mathcal{F}$ with probability

$$\mathbb{P}(\Gamma F(x) = \gamma) = \frac{x^{|\gamma|}}{F(x)|\gamma|!}.$$

In particular, if we condition on a fixed output size n , we get the uniform distribution on \mathcal{F}_n . We describe Boltzmann samplers using an informal pseudo-code notation. Given a specification of the species of structures \mathcal{F} in terms of other species using the operations of sums, products and composition, we obtain a Boltzmann sampler for \mathcal{F} in terms of samplers for the other species involved. The rules for the construction of Boltzmann samplers are summarized

in Table 2. We let $\text{Bern}(p)$ and $\text{Pois}(\lambda)$ denote Bernoulli and Poisson distributed generators.

$\mathcal{F} = \mathcal{A} + \mathcal{B}$	if $\text{Bern}(A(x)/(A(x) + B(x)))$ then return $\Gamma A(x)$ else return $\Gamma B(x)$
$\mathcal{F} = \mathcal{A} \cdot \mathcal{B}$	return $(\Gamma A(x), \Gamma B(x))$ relabeled uniformly at random
$\mathcal{F} = \mathcal{A} \circ \mathcal{B}$	$\gamma \leftarrow \Gamma A(y)$ with $y = B(x)$ for $i = 1$ to $ \gamma $ $\gamma_i \leftarrow \Gamma B(x)$ return $(\gamma, (\gamma_i)_i)$ relabeled uniformly at random
$\mathcal{F} = \text{SET}$	$m \leftarrow \text{Pois}(x)$ return the unique structure of size m

TABLE 2. Rules for the construction of Boltzmann samplers.

Note that if $\mathcal{F} = \mathcal{A}\mu\mathcal{B}$ with $\mathcal{A}, \mathcal{B} \neq 0$, $\mu \in \{+, \cdot, \circ\}$ and $0 < \mathcal{F}(x) < \infty$, then the samplers of \mathcal{A} and \mathcal{B} are almost surely called with valid parameters, since the coefficients of all power-series involved are nonnegative.

Given a combinatorial specification $\mathcal{Y} \simeq \mathcal{H}(\mathcal{X}, \mathcal{Y})$ satisfying the conditions of Theorem 3.1 we may apply the rules above to construct a recursive Boltzmann sampler that is guaranteed to terminate almost surely. This allows us to construct a Boltzmann sampler for block-stable graph classes. More specifically, let \mathcal{C} be a block-stable class of connected graphs such that the radius of convergence ρ of the generating series $C(z)$ is positive. The rooted class \mathcal{C}^\bullet has a combinatorial specification given in (3.3) in terms of the subclass \mathcal{B} of edges and 2-connected graphs. By Lemma 3.2, we know that $y = C^\bullet(\rho)$ and $\lambda = B'(y)$ are finite. We obtain the following sampler which was used before in the study of certain block-stable graph classes, see for example [39].

Corollary 3.5. *Let \mathcal{C} be a block-stable class of connected graphs, $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. The following recursive procedure terminates almost surely and samples according to the Boltzmann distribution for \mathcal{C}^\bullet with parameter ρ .*

$\Gamma\mathcal{C}^\bullet(\rho)$: $\gamma \leftarrow$ a single root vertex
 $m \leftarrow \text{Pois}(\lambda)$
for $k = 1 \dots m$
 $B \leftarrow \Gamma\mathcal{B}'(y)$, drop the labels
 merge γ with the $*$ -vertex of B
 for each non $*$ -vertex v of B
 $C \leftarrow \Gamma\mathcal{C}^\bullet(\rho)$, drop the labels
 merge v with the root of C
return γ relabeled uniformly at random

3.7. Subcritical Graph Classes. Let \mathcal{C} be a block-stable class of connected graphs and \mathcal{B} its subclass of all graphs that are 2-connected or a single edge with its ends. Assume that \mathcal{B} is nonempty and analytic, hence \mathcal{C} is analytic as well by Proposition 3.3. Denote by ρ and R the radii of convergence of the corresponding exponential generating series $C(z)$ and $B(z)$. By Lemma 3.2, we know that ρ , $y = C^\bullet(\rho)$ and $\lambda = B'(y)$ are finite quantities. The following proposition provides a coupling of a Boltzmann-distributed random graph drawn from the

class \mathcal{C} with a Galton-Watson tree. This will play a central role in the proof of the main theorem.

Proposition 3.6. *Let (T, α) denote the enriched tree corresponding to the Boltzmann Sampler $\Gamma\mathcal{C}^\bullet(\rho)$ given in Corollary 3.5. Then the rooted labeled unordered tree T is distributed like the outcome of the following process:*

1. *Draw a Galton-Watson tree with offspring distribution ξ given by the probability generating function $\varphi(z) = \exp(B'(yz) - \lambda)$.*
2. *Distribute labels uniformly at random.*
3. *Discard the ordering on the offspring sets.*

Proof. This follows from the fact that every instance of the sampler $\Gamma\mathcal{C}^\bullet(\rho)$ given in Corollary 3.5 calls itself recursively $|\Gamma(\text{SET} \circ \mathcal{B}')(y)|$ many times. \square

Let ξ denote the offspring distribution given in Proposition 3.6. As discussed above, the rules governing Boltzmann samplers guarantee that the sampler $\Gamma\mathcal{C}^\bullet(\rho)$ terminates almost surely. Hence we have $1 \geq \mathbb{E}[\xi] = \varphi'_\ell(1) = yB''(y) = B'^\bullet(y)$ and in particular $y \leq R$. We define subcriticality depending on whether this inequality is strict.

Definition 3.7. *A block-stable class of connected graphs \mathcal{C} is termed subcritical if $y < R$.*

Prominent examples of subcritical graph classes are trees, outerplanar graphs and series-parallel graphs; the class of planar graphs does not fall into this framework [17, 7], i.e. it satisfies $y = R$. The following lemma was proved in Panagiotou and Steger [39, Lem. 2.8] by analytic methods.

Lemma 3.8. *If $B'^\bullet(R) \geq 1$, then $B'^\bullet(y) = 1$. If $B'^\bullet(R) \leq 1$, then $y = R$. In particular, \mathcal{C} is subcritical if and only if $B'^\bullet(R) > 1$.*

Thus, if $B'^\bullet(R) \geq 1$, then the offspring distribution ξ has expected value 1 and variance

$$\sigma^2 = 1 + B'''(y)y^2 = \mathbb{E}[|\Gamma B'^\bullet(y)|]$$

with $\Gamma B'^\bullet(y)$ denoting a Boltzmann sampler for the class \mathcal{B}^\bullet with parameter y . By Proposition 3.6 the size of the outcome of the sampler $\Gamma\mathcal{C}^\bullet(\rho)$ is distributed like the size of a ξ -Galton-Watson tree. Hence, we may apply Lemma 2.1 to obtain the following result, which was shown in [17] under stronger assumptions.

Corollary 3.9. *Let \mathcal{C} be an analytic block-stable class of graphs, and let ξ be the distribution from Proposition 3.6. Suppose that $B'^\bullet(R) \geq 1$ and $B'''(y) < \infty$, i.e. ξ has finite variance. Let $d = \text{span}(\xi)$. Then, as $n \equiv 1 \pmod d$ tends to infinity,*

$$\mathbb{P}(|\Gamma\mathcal{C}^\bullet(\rho)| = n) \sim \frac{d}{\sqrt{2\pi\mathbb{E}[|\Gamma B'^\bullet(y)|]}} n^{-3/2} \quad \text{and} \quad |\mathcal{C}_n| \sim \frac{yd}{\sqrt{2\pi\mathbb{E}[|\Gamma B'^\bullet(y)|]}} n^{-5/2} \rho^{-n} n!.$$

3.8. Deviation Inequalities. We will make use of the following moderate deviation inequality for one-dimensional random walk found in most textbooks on the subject.

Lemma 3.10. *Let $(X_i)_{i \in \mathbb{N}}$ be family of independent copies of a real-valued random variable X with $\mathbb{E}[X] = 0$. Let $S_n = X_1 + \dots + X_n$. Suppose that there is a $\delta > 0$ such that $\mathbb{E}[e^{\theta X}] < \infty$ for $|\theta| < \delta$. Then there is a $c > 0$ such that for every $1/2 < p < 1$ there is a number N such that for all $n \geq N$ and $0 < \epsilon < 1$*

$$\mathbb{P}(|S_n/n^p| \geq \epsilon) \leq 2 \exp(-c\epsilon^2 n^{2p-1}).$$

4. A SIZE-BIASED RANDOM \mathcal{R} -ENRICHED TREE

Let \mathcal{C} be an analytic block-stable class of connected graphs and $\mathcal{B} \neq 0$ its subclass of graphs that are 2-connected or a single edge. As before we let ρ denote the radius of convergence of the exponential generating series $C(z)$ and set $y = C^\bullet(\rho)$. Recall that by Corollary 3.4 the class \mathcal{C}^\bullet may be identified with the class of \mathcal{R} -enriched trees with $\mathcal{R} := \text{SET} \circ \mathcal{B}'$, i.e. pairs (T, α) with $T \in \mathcal{A}$ a rooted labeled unordered tree and α a function that assigns to each $v \in V(T)$ a (possibly empty) set $\alpha(v)$ of derived blocks whose vertex sets partition the offspring set of the vertex v .

An important ingredient in our forthcoming arguments will be an accurate description of the distribution of the blocks on sufficiently long paths in random graphs from \mathcal{C} . In order to study this distribution we will make use of a special case of a *size-biased* random \mathcal{R} -enriched tree. This construction is based on the size-biased Galton Watson tree introduced in [1] and has several other applications and implications. A thorough study in a more general setting will be presented in Stufler [41].

Recall that $\mathcal{A}_{\mathcal{R}}$ has the decomposition $\mathcal{A}_{\mathcal{R}} \simeq \mathcal{X} \cdot \mathcal{R}(\mathcal{A}_{\mathcal{R}})$. By the rules (3.1) and (3.2) governing operations on species we obtain algebraically

$$\begin{aligned} \mathcal{A}_{\mathcal{R}}^\bullet &\simeq \mathcal{A}_{\mathcal{R}} + \mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_{\mathcal{R}}) \cdot \mathcal{A}_{\mathcal{R}}^\bullet \\ &\simeq \mathcal{A}_{\mathcal{R}} + \mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_{\mathcal{R}}) \cdot \mathcal{A}_{\mathcal{R}} + (\mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_{\mathcal{R}}))^2 \cdot \mathcal{A}_{\mathcal{R}}^\bullet \\ &\dots \\ &\simeq \sum_{\ell \geq 0} (\mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_{\mathcal{R}}))^\ell \mathcal{A}_{\mathcal{R}}. \end{aligned}$$

The above calculation corresponds to taking a direct limit in the category of species, either directly or alternatively by application of Joyal's Implicit Species Theorem. Here $\mathcal{A}_{\mathcal{R}}^{(\ell)} := (\mathcal{X} \cdot \mathcal{R}'(\mathcal{A}_{\mathcal{R}}))^\ell \mathcal{A}_{\mathcal{R}}$ corresponds to the subspecies of all enriched trees (T, α) with a distinguished vertex r such that r has height ℓ in T . It follows from the definition of the Boltzmann distribution that for any pointed enriched tree $(A, r) \in \mathcal{A}_{\mathcal{R}}^{(\ell)}$ we have that

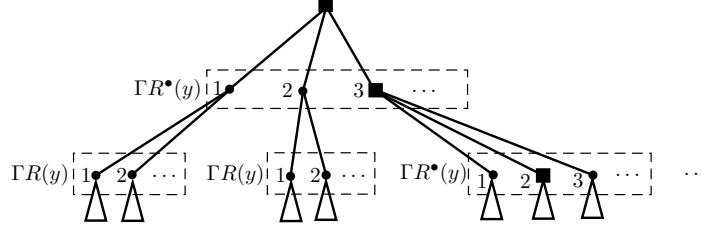
$$(4.1) \quad \mathbb{P} \left(\Gamma A_R^{(\ell)}(\rho) = (A, r) \right) = (\rho R'(y))^{-\ell} \mathbb{P}(\Gamma A_R(\rho) = A).$$

Translating the combinatorial specification for $\mathcal{A}_{\mathcal{R}}^{(\ell)}$ into a Boltzmann sampler $\Gamma A_R^{(\ell)}(\rho)$ yields the following procedure which we call the *size-biased \mathcal{R} -enriched tree* (see also Figure 4). Any vertex is either *normal* or *mutant*, and we start with a single mutant root. Normal vertices have an independent copy of $\Gamma R(y)$ as offspring. Mutant nodes have an independent copy of $\Gamma R^\bullet(y)$ as offspring and the root in the \mathcal{R}^\bullet object is declared mutant, unless it's the ℓ th copy of $\Gamma R^\bullet(y)$. By the theory of recursive Boltzmann samplers obtained from combinatorial specifications this procedure terminates almost surely. The sampler $\Gamma A_R^{(\ell)}(\rho)$ is obtained by additionally distributing labels uniformly at random.

We call the path connecting the inner root with the outer root in an $\mathcal{A}_{\mathcal{R}}^\bullet$ -object the *spine*. Note that the \mathcal{R} -objects along the spine of the random enriched tree $\Gamma A_R^{(\ell)}(\rho)$ are drawn according to ℓ independent copies of $\Gamma R^\bullet(y)$.

In our setting we have that $\mathcal{R} = \text{SET} \circ \mathcal{B}'$, where $\mathcal{B} \neq 0$ denotes the subclass of blocks of the block-stable class \mathcal{C} . Using (3.2) we obtain

$$\mathcal{R}^\bullet \simeq (\text{SET} \circ \mathcal{B}') \cdot \mathcal{B}'^\bullet$$

FIGURE 4. Illustration of the sampler for the size-biased \mathcal{R} -enriched tree.

and the sampler $\Gamma R^\bullet(y)$ is given by independent calls of $\Gamma(\text{SET} \circ \mathcal{B}')(y)$ and $\Gamma \mathcal{B}'^\bullet(y)$. Hence the blocks along the spine are drawn according to ℓ independent copies of $\Gamma \mathcal{B}'^\bullet(y)$.

Equation (4.1) allows us to relate properties of $\Gamma A_R^{(\ell)}(\rho)$ to properties of a uniformly random chosen enriched tree of a given size. We are going to apply the following general lemma in Section 5 in order to show that the blocks along sufficiently long paths in random graphs behave asymptotically like the spine of $\Gamma A_R^{(\ell)}(\rho)$ for a corresponding ℓ .

Lemma 4.1. *Let \mathcal{E} be a property of pointed \mathcal{R} -enriched trees (i.e. a subset of $\mathcal{A}_\mathcal{R}^\bullet$) and let $n \in \mathbb{N}$ be such that $\mathcal{A}_\mathcal{R}[n]$ is nonempty. Consider the function*

$$f : \mathcal{A}_\mathcal{R}[n] \rightarrow \mathbb{R}, \quad A \mapsto \sum_{v \in [n]} \mathbb{1}_{(A,v) \in \mathcal{E}}$$

counting the number of “admissible” outer roots with respect to \mathcal{E} . Let $\mathbf{A}_n \in \mathcal{A}_\mathcal{R}[n]$ be drawn uniformly at random. Then

$$\mathbb{E}[f(\mathbf{A}_n)] = \mathbb{P}(|\Gamma A_R(\rho)| = n)^{-1} \sum_{\ell=0}^{n-1} (\rho R'(y))^\ell \mathbb{P}\left(\Gamma A_R^{(\ell)}(\rho) \text{ has size } n \text{ and satisfies } \mathcal{E}\right).$$

Proof. First, observe that

$$\sum_{v=1}^n \mathbb{P}((\mathbf{A}_n, v) \in \mathcal{E}) = \sum_{\ell=0}^{n-1} \sum_{(A,r) \in \mathcal{E} \cap \mathcal{A}_\mathcal{R}^{(\ell)}[n]} \mathbb{P}(\mathbf{A}_n = A).$$

By (4.1) we have for all $(A, r) \in \mathcal{E} \cap \mathcal{A}_\mathcal{R}^{(\ell)}[n]$ that

$$\mathbb{P}(\Gamma A_R(\rho) = A \mid |\Gamma A_R(\rho)| = n) = (\rho R'(y))^\ell \mathbb{P}\left(\Gamma A_R^{(\ell)}(\rho) = (A, r)\right) \mathbb{P}(|\Gamma A_R(\rho)| = n)^{-1}.$$

This proves the claim. \square

5. CONVERGENCE TOWARDS THE CRT

Let \mathcal{C} be an analytic block-stable class of connected graphs and $\mathcal{B} \neq 0$ its subclass of all graphs that are 2-connected or a single edge. We let $\rho > 0$ denote the radius of convergence of the exponential generating series $C(z)$ and set $y = C^\bullet(\rho)$. As before we identify \mathcal{C}^\bullet with the class $\mathcal{A}_\mathcal{R}$ of \mathcal{R} -enriched trees with $\mathcal{R} = \text{SET} \circ \mathcal{B}'$. By Proposition 3.6 we know that if we draw an \mathcal{R} -enriched tree (\mathbf{T}, α) according to the Boltzmann distribution with parameter ρ , then \mathbf{T} is distributed like a ξ -Galton-Watson tree with $\xi := |\Gamma(\text{SET} \circ \mathcal{B}')(y)|$, relabeling uniformly at random and discarding the ordering on the offspring sets.

Throughout this section let $n \equiv 1 \pmod{\text{span}(\xi)}$ denote a large enough integer such that the probability of a ξ -GWT having size n is positive. Let $C_n \in \mathcal{C}_n$ be drawn uniformly at random and generate $C_n^\bullet \in \mathcal{C}_n^\bullet$ by uniformly choosing a root from $[n]$. We let (T_n, α_n) be the corresponding enriched tree.

For any pointed derived block $B \in \mathcal{B}'^\bullet$ we let $\text{sp}(B) := d_B(*, \text{root})$ denote the length of a shortest path connecting the $*$ -vertex with the root. In this section we prove our main result:

Theorem 5.1. *Let \mathcal{C} be a subcritical class of connected graphs. Then*

$$\frac{\sigma}{2\kappa\sqrt{n}}C_n^\bullet \xrightarrow{(d)} \mathcal{T}_e \quad \text{and} \quad \frac{\sigma}{2\kappa\sqrt{n}}C_n \xrightarrow{(d)} \mathcal{T}_e$$

with respect to the (pointed) Gromov-Hausdorff metric. The constants are given by $\sigma^2 = \mathbb{E}[|B|]$ and $\kappa = \mathbb{E}[\text{sp}(B)]$ with $B \in \mathcal{B}'^\bullet$ a random block drawn according to the Boltzmann distribution with parameter $y = C^\bullet(\rho)$, and in particular $\sigma^2 = 1 + B'''(y)y^2$.

As a consequence we obtain the limit distributions for the height and diameter of C_n^\bullet .

Corollary 5.2. *Let \mathcal{C} be a subcritical class of connected graphs. Then the rescaled height $\frac{\sigma}{2\kappa\sqrt{n}}H(C_n^\bullet)$ and diameter $\frac{\sigma}{2\kappa\sqrt{n}}D(C_n)$ converge in distribution to $H(\mathcal{T}_e)$ and $D(\mathcal{T}_e)$, i.e. for all $x > 0$, as n tends to infinity*

$$\begin{aligned} \mathbb{P}\left(H(C_n^\bullet) > \frac{2\kappa\sqrt{n}}{\sigma}x\right) &\rightarrow 2 \sum_{k=1}^{\infty} (4k^2x^2 - 1) \exp(-2k^2x^2), \\ \mathbb{P}\left(D(C_n) > \frac{2\kappa\sqrt{n}}{\sigma}x\right) &\rightarrow \sum_{k=1}^{\infty} (k^2 - 1) \left(\frac{2}{3}k^4x^4 - 4k^2x^2 + 2\right) \exp(-k^2x^2/2). \end{aligned}$$

Moreover, all moments converge as well. In particular

$$\mathbb{E}[D(C_n)] \sim \frac{2^{5/2}\kappa}{3\sigma} \sqrt{\pi n} \sim \frac{4}{3} \mathbb{E}[H(C_n^\bullet)].$$

Expressions for arbitrarily high moments are given in (2.6) and (2.8).

Proof. The limiting distributions are given in (2.5) and (2.7). In order to show convergence of the moments we argue that the rescaled height and diameter are bounded in the space L^p for all $1 < p < \infty$. This follows for example from the subgaussian tail-bounds of Theorem 6.1 given in Section 6 below (note that the proof of Theorem 6.1 does not depend on the results in this section). \square

In the following we are going to prove Theorem 5.1. The idea is to show that the pointed Gromov-Hausdorff distance of C_n^\bullet and κT_n is small with high probability and use the convergence of T_n towards a multiple of the CRT \mathcal{T}_e .

Definition 5.3. *Let $C \in \mathcal{C}$. For any $x, y \in V(C)$ set $\bar{d}_C(x, y) := d_T(x, y)$ with (T, α) the enriched tree corresponding to (C, x) , i.e. C rooted at the vertex x .*

Less formally speaking, $\bar{d}_C(x, y)$ denotes the minimum number of blocks required to cover the edges of a shortest path linking x and y . It takes a moment to see that if (T, α) corresponds to the rooted graph (C, z) , then $\bar{d}_C(x, y) \leq d_T(x, y) \leq \bar{d}_C(x, y) + 1$ for all $x, y \in V(C)$. In particular, \bar{d}_C is a metric: the triangle inequality holds since d_T is a metric and

$$\bar{d}_C(x, y) \leq d_T(x, y) \leq d_T(x, z) + d_T(y, z) = \bar{d}_C(x, z) + \bar{d}_C(y, z).$$

In the following lemma we apply the results on pointed enriched trees of Section 4.

Lemma 5.4. *Let \mathcal{C} be a subcritical class of connected graphs and set $\kappa = \mathbb{E}[\text{sp}(\Gamma B'^{\bullet}(y))]$. Then for all $s > 1$ and $0 < \epsilon < 1/2$ with $2\epsilon s > 1$ we have with high probability that all $x, y \in V(\mathbf{C}_n)$ with $\bar{d}_{\mathbf{C}_n}(x, y) \geq \log^s(n)$ satisfy $|d_{\mathbf{C}_n}(x, y) - \kappa \bar{d}_{\mathbf{C}_n}(x, y)| \leq \bar{d}_{\mathbf{C}_n}(x, y)^{1/2+\epsilon}$.*

Proof. We denote $L_n = \log^s(n)$ and $t_\ell = \ell^{1/2+\epsilon}$. Let $\mathcal{E} \subset \mathcal{A}_{\mathcal{R}}^{\bullet} \simeq \mathcal{C}^{\bullet\bullet}$ with $\mathcal{R} = \text{SET} \circ \mathcal{B}'$ denote the set of all bipointed graphs or pointed enriched trees $((C, x), y) \simeq ((T, \alpha), y)$, where we call x the *inner root* and y the *outer root*, such that

$$d_T(x, y) \geq L_{|T|} \quad \text{and} \quad |d_C(x, y) - \kappa d_T(x, y)| > t_{d_T(x, y)}.$$

We will bound the probability that there exist vertices x and y with $((\mathbf{C}_n, x), y) \in \mathcal{E}$. First observe that

$$\sum_{x, y \in [n]} \mathbb{P}(((\mathbf{C}_n, x), y) \in \mathcal{E}) = \sum_{((C, x), y) \in \mathcal{E}} \mathbb{P}(\mathbf{C}_n = C) = n \sum_{y=1}^n \mathbb{P}((\mathbf{C}_n^{\bullet}, y) \in \mathcal{E}).$$

By assumption we may apply Corollary 3.9 to obtain $\mathbb{P}(|\Gamma \mathcal{C}^{\bullet}(\rho)| = n) = \Theta(n^{-3/2})$. Moreover, Lemma 3.8 asserts that $B'^{\bullet}(y) = 1$ and thus, with Lemma 3.2

$$\rho R'(y) = \rho B''(y) e^{B'(y)} = y B''(y) = 1.$$

Hence, by applying Lemma 4.1 we obtain that

$$\mathbb{P}(((\mathbf{C}_n, x), y) \in \mathcal{E} \text{ for some } x, y) \leq O(n^{5/2}) \sum_{\ell=L_n}^{n-1} \mathbb{P}\left(\Gamma A_R^{(\ell)}(\rho) \text{ has size } n \text{ and satisfies } \mathcal{E}\right).$$

The height of the outer root in the bipointed graph corresponding to $\Gamma A_R^{(\ell)}(\rho)$ is distributed like the sum of ℓ independent random variables, each distributed like the distance of the $*$ -vertex and the root in the corresponding derived block of $\Gamma(\text{SET} \circ \mathcal{B}')^{\bullet}(y)$. Since $(\text{SET} \circ \mathcal{B}')^{\bullet} \simeq (\text{SET} \circ \mathcal{B}') \cdot \mathcal{B}'^{\bullet}$, these variables are actually $\text{sp}(\Gamma B'^{\bullet}(y))$ -distributed. Hence

$$\mathbb{P}\left(\Gamma A_R^{(\ell)}(\rho) \in \mathcal{E}, |\Gamma A_R^{(\ell)}(\rho)| = n\right) \leq \mathbb{P}(|\eta_1 + \dots + \eta_\ell - \ell \mathbb{E}[\eta_1]| > t_\ell)$$

with $(\eta_i)_i$ i.i.d. copies of $\eta := \text{sp}(\Gamma B'^{\bullet}(y))$. Clearly we have that $\eta \leq |\Gamma B'^{\bullet}(y)|$. Since \mathcal{C} is subcritical it follows that there is a constant $\delta > 0$ such that $\mathbb{E}[e^{\eta^\theta}] < \infty$ for all θ with $|\theta| \leq \delta$. Hence we may apply the standard moderate deviation inequality for one-dimensional random walk stated in Lemma 3.10 to obtain for some constant $c > 0$

$$\mathbb{P}(((\mathbf{C}_n, x), y) \in \mathcal{E} \text{ for some } x, y) \leq O(n^{7/2}) \exp(-c(\log n)^{2s\epsilon}) = o(1).$$

□

It remains to clarify what happens if $\bar{d}_{\mathbf{C}_n}$ is small. We prove the following statement for random graphs from block-stable classes that are not necessarily subcritical.

Proposition 5.5. *Let \mathcal{C} be a block-stable class of connected graphs. Suppose that $B'^{\bullet}(y) = 1$ and the offspring distribution ξ has finite second moment, i.e. $B'''(y) < \infty$. Let $\text{lb}(\mathbf{C}_n)$ denote the size of the largest block in \mathbf{C}_n ,*

- (1) *For any $x, y \in \mathbf{C}_n$ we have $d_{\mathbf{C}_n}(x, y) \leq \bar{d}_{\mathbf{C}_n}(x, y) \text{lb}(\mathbf{C}_n)$.*
- (2) *If the offspring distribution ξ is bounded, then so is $\text{lb}(\mathbf{C}_n)$. Otherwise, for any sequence K_n we have $\mathbb{P}(\text{lb}(\mathbf{C}_n) \geq K_n) = O(n) \mathbb{P}(\xi \geq K_n)$.*

Proof. We have that $d_{C_n} \leq \bar{d}_{C_n}(\text{lb}(C_n) - 1)$ and $\text{lb}(C_n) = \text{lb}(C_n^\bullet) \leq \Delta(T_n) + 1$ with $\Delta(T_n)$ denoting the largest outdegree. Recall that $\Delta(T_n)$ is distributed like the maximum degree of a ξ -Galton-Watson tree conditioned to have n vertices. By assumption, the offspring distribution ξ has expected value $\mathbb{E}[\xi] = B'^\bullet(y) = 1$ and finite variance.

If ξ is bounded then so is the largest outdegree of T_n . Otherwise, as argued in the proof of [27, Eq. (19.20)], for any sequence K_n

$$(5.1) \quad \mathbb{P}(\Delta(T_n) \geq K_n) \leq (1 + o(1))n\mathbb{P}(\xi \geq K_n).$$

Applying (5.1) yields $\mathbb{P}(\Delta(T_n) \geq K_n) \leq (1 + o(1))n\mathbb{P}(\xi \geq K_n)$ for any sequence K_n . \square

Note that if \mathcal{C} is subcritical then this implies that $\text{lb}(C_n) = O(\log n)$ with high probability: the definition of the Boltzmann model and the fact that y is smaller than the radius of convergence of $B(z)$ guarantee that there is a constant $\beta < 1$ such that

$$\mathbb{P}(\xi = k) = \mathbb{P}(|\Gamma(\text{SET} \circ B')(y)| = k) = O(\beta^k).$$

Combined with the bounds of Lemma 5.4 this yields the following concentration result.

Corollary 5.6. *Let \mathcal{C} be a subcritical class of connected graphs. Then for all $s > 1$ and $0 < \epsilon < 1/2$ with $2\epsilon s > 1$ we have with high probability that for all vertices $x, y \in V(C_n)$*

$$|d_{C_n}(x, y) - \kappa \bar{d}_{C_n}(x, y)| \leq \bar{d}_{C_n}(x, y)^{1/2+\epsilon} + O(\log^{s+1}(n)).$$

We may now prove the main theorem.

Proof of Theorem 5.1. Recall that $\bar{d}_{C_n} \leq d_{T_n} \leq \bar{d}_{C_n} + 1$. By Corollary 5.6, Proposition 2.3, and considering the distortion of the identity map as correspondence between the vertices of T_n and C_n^\bullet , it follows that with high probability

$$d_{\text{GH}}(C_n^\bullet/(\kappa\sqrt{n}), T_n/\sqrt{n}) \leq D(T_n)^{3/4}/\sqrt{n} + o(1).$$

Using the tail bounds (2.1) for the diameter $D(T_n)$ we obtain that $d_{\text{GH}}(C_n^\bullet/(\kappa\sqrt{n}), T_n/\sqrt{n})$ converges in probability to zero. Recall that the variance of the offspring distribution ξ is given by $\sigma^2 = \mathbb{E}[|\Gamma B'^\bullet(y)|]$. By Theorem 2.5 we have that $\frac{\sigma}{2\sqrt{n}}T_n \xrightarrow{(d)} \mathcal{T}_\epsilon$ and thus $\frac{\sigma}{2\kappa\sqrt{n}}C_n^\bullet \xrightarrow{(d)} \mathcal{T}_\epsilon$. \square

6. SUBGAUSSIAN TAIL BOUNDS FOR THE HEIGHT AND DIAMETER

In this section we prove subgaussian tail bounds for the height and diameter of the random graphs C_n^\bullet and C_n . Our proof builds on results obtained in [1]. Recall that (T_n, α_n) denotes the enriched tree corresponding to the graph C_n^\bullet and that T_n has a natural coupling with a ξ -Galton-Watson conditioned on having size n , see Proposition 3.6. With (slight) abuse of notation we also write T_n for the conditioned ξ -Galton-Watson tree within this section. We prove the following statement for random graphs from block-stable classes that are not necessarily subcritical.

Theorem 6.1. *Let \mathcal{C} be a block-stable class of connected graphs. Suppose that \mathcal{C} satisfies $B'^\bullet(y) = 1$ and the offspring distribution ξ has finite variance, i.e. $B'''(y) < 1$. Then there are $C, c > 0$ such that for all $n, x \geq 0$*

$$\mathbb{P}(D(C_n) \geq x) \leq C \exp(-cx^2/n) \quad \text{and} \quad \mathbb{P}(H(C_n^\bullet) \geq x) \leq C \exp(-cx^2/n).$$

As $H(T_n) \leq H(C_n^\bullet)$, Inequality (2.1) also yields a lower tail bound for the height of C_n^\bullet .

Corollary 6.2. *There are constants $C, c > 0$ such that for all $x \geq 0$ and n*

$$\mathbb{P}(H(\mathbf{C}_n^\bullet) \leq x) \leq C \exp(-c(n-2)/x^2).$$

As a main ingredient in our proof we consider the *lexicographic depth-first-search (DFS)* of the plane tree T_n by labeling the vertices in the usual way (as a subtree of the Ulam-Harris tree) by finite sequences of integers and listing them in lexicographic order v_0, v_1, \dots, v_{n-1} . The search keeps a queue of Q_i^d nodes with $Q_0^d = 1$ and the recursion

$$Q_i^d = Q_{i-1}^d - 1 + d_{\mathsf{T}_n^p}^+(v_{i-1}).$$

The mirror-image of T_n is obtain by reversing the ordering on each offspring set and the *reverse DFS* Q_i^r is defined as the DFS of the mirror-image. Then $(Q_i^d)_{0 \leq i \leq n}$ and $(Q_i^r)_{0 \leq i \leq n}$ are identically distributed and satisfy the following bound given in [1, Ineq. (4.4)]:

$$(6.1) \quad \mathbb{P}\left(\max_j Q_j \geq x\right) \leq C \exp(-cx^2/n) \quad \text{where } Q_j \in \{Q_j^d, Q_j^r\}$$

with $C, c > 0$ denoting some constants that do not depend on x or n .

Proof of Theorem 6.1. Since $D(\mathbf{C}_n) \leq 2H(\mathbf{C}_n^\bullet)$ it suffices to show the bound for the height. Let $h \geq 0$. If $H(\mathbf{C}_n^\bullet) \geq h$ then there exists a vertex whose height equals h . Consequently, we may estimate $\mathbb{P}(H(\mathbf{C}_n^\bullet) \geq h) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2)$ with \mathcal{E}_1 (resp. \mathcal{E}_2) denoting the event that there is a vertex v such that $h_{\mathbf{C}_n^\bullet}(v) = h$ and $h_{\mathsf{T}_n}(v) \geq h/2$ (resp. $h_{\mathsf{T}_n}(v) \leq h/2$). By the tail bound (2.1) for the height of Galton-Watson trees we obtain

$$\mathbb{P}(\mathcal{E}_1) \leq \mathbb{P}(H(\mathsf{T}_n) \geq h) \leq C_2 \exp(-c_2 h^2/(4n))$$

for some constants $C_2, c_2 > 0$. In order to bound $\mathbb{P}(\mathcal{E}_2)$ suppose that there is a vertex v with height $h_{\mathbf{C}_n^\bullet}(v) = h$ and $h_{\mathsf{T}_n}(v) \leq h/2$. If a is a vertex of T_n and b one of its offspring, then $d_{\mathbf{C}_n^\bullet}(a, b) \leq d_{\mathsf{T}_n}^+(a)$. Hence

$$\sum_{u \succ v} d_{\mathsf{T}_n}^+(u) \geq h_{\mathbf{C}_n^\bullet}(v) = h$$

with the sum index u ranging over all ancestors of v . Consider the lexicographic depth-first-search $(Q_i^d)_i$ and reverse depth-first-search $(Q_i^r)_i$ of T_n^p . Let j (resp. k) denote the index corresponding to the vertex v in the lexicographic (resp. reverse lexicographic) order. It follows from the definition of the queues that if \mathcal{E}_2 occurs

$$Q_j^d + Q_k^r = 2 + \sum_{u \succ v} d_{\mathsf{T}_n^p}^+(u) - h_{\mathsf{T}_n}(v) \geq h/2$$

and hence $\max(Q_j^d, Q_k^r) \geq h/4$. Since Q_j^d and Q_k^r are identically distributed it follows by (6.1) that

$$\begin{aligned} \mathbb{P}(\mathcal{E}_2) &\leq \mathbb{P}\left(\max_i (Q_i^d) \geq h/4\right) + \mathbb{P}\left(\max_i (Q_i^r) \geq h/4\right) \\ &\leq 2\mathbb{P}\left(\max_i (Q_i^d) \geq h/4\right) \\ &\leq 2C \exp(-ch^2/(16n)). \end{aligned}$$

This concludes the proof. \square

7. EXTENSIONS

In the following we use the notation from Section 5.

7.1. First Passage Percolation. Let $\omega > 0$ be a given random variable such that there is a $\delta > 0$ with $\mathbb{E}[e^{\theta\omega}] < \infty$ for all θ with $|\theta| \leq \delta$. For any graph G we may consider the random graph \hat{G} obtained by assigning to each edge $e \in E(G)$ a weight ω_e that is an independent copy of ω . The $d_{\hat{G}}$ -distance of two vertices a and b is then given by

$$d_{\hat{G}}(a, b) = \inf \left\{ \sum_{e \in E(P)} \omega_e \mid P \text{ a path connecting } a \text{ and } b \text{ in } G \right\}.$$

Let $\hat{\kappa} := \mathbb{E}[\text{sp}(\hat{\mathbf{B}})]$ with \mathbf{B} drawn according to the Boltzmann sampler $\Gamma B^\bullet(y)$ and $\text{sp}(\hat{\mathbf{B}})$ denoting the $d_{\hat{\mathbf{B}}}$ -distance from the $*$ -vertex to the root vertex.

Theorem 7.1. *Let \mathcal{C} be a subcritical class of connected graphs. We have that*

$$\frac{\sigma}{2\hat{\kappa}\sqrt{n}} \hat{\mathbf{C}}_n \xrightarrow{(d)} \mathcal{T}_e \quad \text{and} \quad \frac{\sigma}{2\hat{\kappa}\sqrt{n}} \hat{\mathbf{C}}_n \xrightarrow{(d)} \mathcal{T}_e$$

with respect to the (pointed) Gromov-Hausdorff metric.

Proof. For any n let K_n denote the complete graph with n vertices. The idea is to generate $\hat{\mathbf{C}}_n$ by drawing \mathbf{C}_n and \hat{K}_n independently and assign the weights via the inclusion $E(\mathbf{C}_n) \subset E(K_n)$. By considering subsets $\mathcal{E} \subset \mathcal{C}^\bullet \times \mathbb{R}^{\cup_n E(K_n)}$ we may easily prove a weighted version of Lemma 4.1, i.e. the probability that the random pair $(\mathbf{C}_n^\bullet, \hat{K}_n)$ has some property \mathcal{E} is bounded by

$$O(n^{5/2}) \sum_{\ell=0}^{n-1} \mathbb{P} \left(|\Gamma \mathbf{C}^{\bullet(\ell)}(\rho)| = n, (\Gamma \mathbf{C}^{\bullet(\ell)}(\rho), \hat{K}_n) \in \mathcal{E} \right).$$

This implies that the blocks along sufficiently long paths in the random graphs $\hat{\mathbf{C}}_n$ behave like independent copies of the weighted block $\hat{\mathbf{B}}$ with \mathbf{B} drawn according to the Boltzmann sampler $\Gamma B^\bullet(y)$. Hence, weighted versions of Lemma 5.4 and Proposition 5.5 may be deduced analogously to their original proofs with $\hat{\kappa}$ replacing κ and only minor modifications otherwise. Thus the scaling limit follows in the same fashion. \square

7.2. Random Graphs given by Their Connected Components. We study the case of an arbitrary graph consisting of a set of connected components. Let $\mathcal{G} \simeq \text{SET} \circ \mathcal{C}$ denote a subcritical graph class given by its subclass \mathcal{C} of connected graphs. For simplicity we are going to assume that all trees belong to the class \mathcal{C} .

Consider the uniform random graph $\mathbf{G}_n \in \mathcal{G}_n$. Of course we cannot expect \mathbf{G}_n to converge to the Continuum Random Tree since may be disconnected with a probability that is bounded away from zero. Instead we study a uniformly chosen component \mathbf{H}_n of maximal size. We are going to show that \mathbf{H}_n/\sqrt{n} converges to a multiple of the CRT.

Theorem 7.2. *Suppose \mathcal{C} is subcritical class of connected graphs containing all trees. Then $\frac{\sigma}{2\kappa\sqrt{n}} \mathbf{H}_n \xrightarrow{(d)} \mathcal{T}_e$ with respect to the Gromov-Hausdorff metric, where σ, κ are as in Theorem 5.1.*

We are going to use the known fact that with high probability the random graph \mathbf{G}_n has a unique giant component with size $n + O_p(1)$. This follows for example from [37, Thm. 6.4].

Lemma 7.3. *If \mathcal{C} contains all trees, then the size of a largest component satisfies $|\mathbf{H}_n| = n + O_p(1)$.*

Proof of Theorem 7.2. Let $f : \mathbb{K} \rightarrow \mathbb{R}$ be a bounded Lipschitz-continuous function defined on the space of isometry classes of compact metric spaces. We will show that $\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathbf{H}_n)] \rightarrow \mathbb{E}[f(\mathcal{T}_e)]$ as n tends to infinity. Set $\Omega_n := \log n$. By Lemma 7.3 we know that $|\mathbf{H}_n| = n + O_p(1)$. Hence with high probability we have that $n - |\mathbf{H}_n| \leq \Omega_n$ and thus

$$\mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathbf{H}_n\right)\right] = o(1) + \sum_{0 \leq k \leq \Omega_n} \mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathbf{H}_n\right) \mid |\mathbf{H}_n| = n - k\right] \mathbb{P}(|\mathbf{H}_n| = n - k).$$

The conditional distribution of \mathbf{G}_n given the sizes $(s_i)_i$ of its components is given by choosing components $K_i \in \mathcal{C}[s_i]$ independently uniformly at random and distributing labels uniformly at random. In particular, as a metric space, \mathbf{H}_n conditioned on $|\mathbf{H}_n| = n - k$ is distributed like the uniform random graph \mathbf{C}_{n-k} . Thus, given $\epsilon > 0$ we have for n sufficiently large by Lipschitz-continuity

$$\mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathbf{H}_n\right) \mid |\mathbf{H}_n| = n - k\right] = \mathbb{E}\left[f\left(\frac{\sigma}{2\kappa\sqrt{n}}\mathbf{C}_{n-k}\right)\right] \in \mathbb{E}[f(\mathcal{T}_e)] \pm \epsilon$$

for all $0 \leq k \leq \Omega_n$. Thus $|\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathbf{H}_n)] - \mathbb{E}[f(\mathcal{T}_e)]| \leq \epsilon$ for sufficiently large n . Since $\epsilon > 0$ was chosen arbitrarily it follows that $\mathbb{E}[f(\frac{\sigma}{2\kappa\sqrt{n}}\mathbf{H}_n)] \rightarrow \mathbb{E}[f(\mathcal{T}_e)]$ as n tends to infinity. \square

8. THE SCALING FACTOR OF SPECIFIC CLASSES

In this section we apply our main results to several specific examples of subcritical graph classes. The notation that will be fixed throughout this section is as follows: \mathcal{C} denotes a subcritical class of connected graphs and \mathcal{B} its subclass of 2-connected graphs and edges. The radius of convergence of $C(z)$ is denoted by ρ . The constant $y = \mathcal{C}^\bullet(\rho)$ is the unique positive solution of the equation

$$yB''(y) = 1.$$

By Lemma 3.2 this determines $\rho = y \exp(-B'(y))$. Moreover, we set

$$\kappa = \mathbb{E}[\text{sp}(\Gamma B'^\bullet(y))],$$

i.e. the expected distance from the $*$ -vertex to the root in a random block chosen according to the Boltzmann distribution with parameter y . We call κ the *scaling factor* for \mathcal{C} . The offspring distribution ξ of the random tree corresponding to the sampler $\Gamma \mathcal{C}^\bullet(y)$ has probability generating function $\varphi(z) = \exp(B'(yz) - \lambda)$ with $\lambda = B'(y)$, see Proposition 3.6. Its variance is given by

$$\sigma^2 = 1 + B'''(y)y^2 = \mathbb{E}[|\Gamma B'^\bullet(y)|].$$

We let d denote the span of the offspring distribution. By applying Corollary 5.2 we obtain

$$\mathbb{E}[H(\mathbf{C}_n^\bullet)]/\sqrt{n} \rightarrow \kappa\sqrt{2\pi/\sigma^2} =: H \quad \text{as } n \rightarrow \infty \text{ with } n \equiv 1 \pmod{d}$$

with $\mathbf{C}_n^\bullet \in \mathcal{C}_n^\bullet$ drawn uniformly at random. We call H the *expected rescaled height*. Moreover, by applying Corollary 3.9 we may assume that

$$|\mathcal{C}_n| \sim c n^{-5/2} \rho_C^{-n} n! \quad \text{as } n \rightarrow \infty \text{ with } n \equiv 1 \pmod{d}$$

with $c = yd/\sqrt{2\pi\sigma^2}$. In this section we derive analytical expressions for the relevant constants $\kappa, H, c, \rho, y, \lambda, \sigma^2$ for several graph classes; Table 3 provides numerical approximations. For a set of graphs M , we denote by $\text{Forb}(M)$ the class of all connected graphs that contain none

Graph Class	κ	H	c	ρ	y	λ	σ^2
Trees = $\text{Forb}(C_3)$	1	2.50662	0.39894	0.36787	1	1	1
$\text{Forb}(C_4)$	1	2.13226	0.20973	0.23618	0.27520	0.80901	1.38196
$\text{Forb}(C_5)$	1.10355	1.88657	0.10987	0.06290	0.40384	1.85945	2.14989
Cacti Graphs	1.20297	1.99021	0.12014	0.23874	0.45631	0.64779	2.29559
Outerplanar Graphs	5.08418	1.30501	0.00697	0.13659	0.17076	0.22327	95.3658

TABLE 3. Numerical approximations of constants for examples of subcritical classes of connected graphs.

of the graphs in M as a topological minor; if M contains only 2-connected graphs, then it is easy to see that $\text{Forb}(M)$ is block-stable, cf. [16]. For $n \geq 3$ we denote by C_n a graph that is isomorphic to a cycle with n vertices.

8.1. Trees. Let \mathcal{C} be the class of trees, i.e. \mathcal{B} consists only of the graph K_2 . It is easy to see that the offspring distribution follows a Poisson distribution with parameter one. We immediately obtain:

Proposition 8.1. *For the class of trees we have $\kappa = 1$ and $\sigma^2 = 1$.*

8.2. $\text{Forb}(C_4)$. Let \mathcal{C} denote the connected graphs of the class $\text{Forb}(C_4)$. Then each block is either isomorphic to K_2 or K_3 . Hence $B(z) = z^2/2 + z^3/6$. Moreover, for any $B \in \mathcal{B}$ and any two distinct vertices in B their distance is one. A simple computation then yields:

Proposition 8.2. *For the class $\text{Forb}(C_4)$ we have $\kappa = 1$ and $\sigma^2 = (5 - \sqrt{5})/2$.*

8.3. $\text{Forb}(C_5)$. Recall that the class $\text{Forb}(C_5)$ consists of all graphs that do not contain a cycle with five vertices as a topological minor. Hence, a graph belongs to this class if and only if it contains no cycle of length at least five as subgraph.

Proposition 8.3. *For the class $\text{Forb}(C_5)$ the constant y is the unique positive solution to $zB''(z) = 1$, where B' is given in (8.3). Moreover, we have*

$$\kappa = (2y^2 + 4y + 3)ye^y - (3y^2 + 12y + 4)y/2 \approx 1.10355.$$

and $\sigma^2 = 1 + B'''(y)y^2 \approx 2.14989$.

Before proving Proposition 8.3 we identify the unlabeled blocks of this class. This result (among extensions to $\text{Forb}(C_6)$ and $\text{Forb}(C_7)$) was given by Giménez, Mitsche and Noy [24].

Proposition 8.4. *The unlabeled blocks of the class $\text{Forb}(C_5)$ are given by*

$$(8.1) \quad K_2, K_4, (K_{2,m})_{m \geq 1}, (K_{2,m}^+)_{m \geq 2}.$$

Here K_n denotes the complete graph and $K_{m,n}$ the complete bipartite graph with bipartition $\{A_m, B_n\}$. The graph $K_{2,n}^+$ is obtained from $K_{2,n}$ by adding an additional edge between the two vertices from A_2 .

Proof. We may verify (8.1) by considering the standard decomposition of 2-connected graphs: an arbitrary graph G is 2-connected if and only if it can be constructed from a cycle by adding H -paths to already constructed graphs H [16]. If $G \in \text{Forb}(C_5)$, then so do all the graphs along its decomposition. In particular we must start with a triangle or a square. Since every

edge of a 2-connected graph lies on a cycle, we may only add paths of length at most two in each step. In particular, for $m \geq 3$ a $K_{2,m}$ may only become a $K_{2,m}^+$ or $K_{2,m+1}$, and a $K_{2,m}^+$ may only become a $K_{2,m+1}^+$. Thus (8.1) follows by induction on the number of vertices. \square

Proof of Proposition 8.3. With foresight, we use the decomposition

$$(8.2) \quad \mathcal{B} = \mathcal{S} + \mathcal{H} + \mathcal{P}$$

with the classes of labeled graphs \mathcal{S} , \mathcal{H} and \mathcal{P} defined by their sets of unlabeled graphs $\tilde{\mathcal{S}} = \{K_2, K_3, K_4, C_4\}$, $\tilde{\mathcal{H}} = \{K_{2,m} \mid m \geq 3\}$ and $\tilde{\mathcal{P}} = \{K_{2,m}^+ \mid m \geq 2\}$. Any unlabeled graph from \mathcal{H} or \mathcal{P} with n vertices has exactly $\binom{n}{2}$ different labelings, since any labeling is determined by the choice of the two unique vertices with degree at least three. Hence

$$S(x) = x^2/2 + x^3/6 + x^4/6, \quad H(x) = \sum_{n \geq 5} \binom{n}{2} \frac{x^n}{n!} \quad \text{and} \quad P(x) = \sum_{n \geq 4} \binom{n}{2} \frac{x^n}{n!}$$

and thus

$$(8.3) \quad B'(x) = x(x+2)e^x - x(15x + 2x^2 + 6)/6.$$

Solving the equation $B'^{\bullet}(y) = 1$ yields

$$y \approx 0.40384.$$

First, let $H_n \in \mathcal{H}'^{\bullet}_n$ with $n \geq 4$ be drawn uniformly at random. We say that a vertex lies on the left if it has degree at least three, otherwise we say it lies on the right. There are $n \binom{n+1}{2}$ graphs in the class \mathcal{H}'^{\bullet}_n and precisely n^2 of those have the property that the $*$ -vertex lies on the left. The distance of the root and the $*$ -vertex equals two if they lie on the same side and one otherwise. Hence

$$\mathbb{E}[\text{sp}(H_n)] = \frac{n}{\binom{n+1}{2}} \left(\frac{1}{n} \cdot 2 + \frac{n-1}{n} \cdot 1 \right) + \left(1 - \frac{n}{\binom{n+1}{2}} \right) \left(\frac{2}{n} \cdot 1 + \frac{n-2}{n} \cdot 2 \right).$$

Let $P_n \in \mathcal{P}'^{\bullet}_n$ with $n \geq 3$ and $S_n \in \mathcal{S}_n$ with $n = 1, 2, 3$ be drawn uniformly at random. Analogously to the above calculation we obtain

$$\mathbb{E}[\text{sp}(P_n)] = \frac{n}{\binom{n+1}{2}} 1 + \left(1 - \frac{n}{\binom{n+1}{2}} \right) \left(\frac{2}{n} \cdot 1 + \frac{n-2}{n} \cdot 2 \right)$$

and

$$\mathbb{E}[\text{sp}(S_1)] = \mathbb{E}[\text{sp}(S_2)] = 1, \quad \mathbb{E}[\text{sp}(S_3)] = \frac{1}{4} \cdot 1 + \frac{3}{4} \left(\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 2 \right) = \frac{5}{4}.$$

Since $B'^{\bullet}(y) = 1$ we have for any class $\mathcal{F} \in \{\mathcal{S}'^{\bullet}, \mathcal{H}'^{\bullet}, \mathcal{P}'^{\bullet}\}$ that

$$\mathbb{E}[\text{sp}(\Gamma B'^{\bullet}(y)), \Gamma B'^{\bullet}(y) \in \mathcal{F}] = \sum_n ([z^n] F(yz)) \mathbb{E}[\text{sp}(F_n)].$$

Summing up yields

$$\mathbb{E}[\text{sp}(\Gamma B'^{\bullet}(y))] = (2y^2 + 4y + 3)ye^y - (3y^2 + 12y + 4)y/2 \approx 1.10355.$$

\square

8.4. Cacti Graphs. A cactus graph is a graph in which each edge is contained in at most one cycle. Equivalently, the class of cacti graphs is the block-stable class of graphs where every block is either an edge or a cycle. In the following \mathcal{C} denotes the class of cacti graphs.

Proposition 8.5. *For the class of cacti graphs the constant y is the unique positive solution to $zB''(z) = 1$, where B' is given in (8.4). Moreover, we have*

$$\kappa = \frac{y^4 - 2y^3 + 2y - 2}{(y^2 - 2y + 2)(1 + y)(y - 1)} \approx 1.20297.$$

and $\sigma^2 = 1 + B'''(y)y^2 \approx 2.29559$.

Proof. By counting the number of labelings of a cycle, we obtain $|\mathcal{B}'_n| = n!/2$ for $n \geq 2$. It follows that

$$(8.4) \quad B'(z) = z + \frac{z^2}{2(1 - z)}$$

and hence $B'^{\bullet}(z) = z + \frac{1}{2} \sum_{n \geq 2} n z^n = \frac{z^3 - 2z^2 + 2z}{2(z-1)^2}$. Solving the equation $B'^{\bullet}(y) = 1$ yields

$$y = -\frac{1}{3}(17 + 3\sqrt{33})^{1/3} + \frac{2}{3}(17 + 3\sqrt{33})^{-1/3} + \frac{4}{3} \approx 0.45631.$$

Let $\Gamma B'^{\bullet}(y)$ denote a Boltzmann-sampler for the class \mathcal{B}'^{\bullet} with parameter y and for any $n \geq 1$ let $B_n \in \mathcal{B}'_n$ be drawn uniformly at random. Since $B'^{\bullet}(y) = 1$, it follows that

$$\kappa = \mathbb{E}[\mathbb{E}[\text{sp}(\Gamma B'^{\bullet}(y)) \mid |\Gamma B'^{\bullet}(y)|]] = \sum_{n \geq 1} \text{sp}(B_n)[z^n] B'^{\bullet}(yz) = \text{sp}(B_1)y + \frac{1}{2} \sum_{n \geq 2} \text{sp}(B_n)ny^n.$$

Clearly $\text{sp}(B_1) = 1$ and for $n \geq 2$ we have that $\text{sp}(B_n)$ is distributed like the distance from the $*$ -vertex to a uniformly at random chosen root from $[n]$ in the cycle $(*, 1, 2, \dots, n)$. Hence

$$\text{sp}(B_n) = \begin{cases} \frac{2}{n} \sum_{i=1}^{n/2} i = \frac{n+2}{4}, & n \text{ is even} \\ \frac{n+1}{2n} + \frac{2}{n} \sum_{i=1}^{(n-1)/2} i = \frac{(n+1)^2}{4n}, & n \text{ is odd} \end{cases}.$$

Summing up over all possible values of n yields the claimed expression for κ . \square

8.5. Outerplanar Graphs. An outerplanar graph is a planar graph that can be embedded in the plane in such a way that every vertex lies on the boundary of the outer face. Let \mathcal{C} denote the class of connected outerplanar graphs and \mathcal{B} the subclass consisting of single edges or 2-connected outerplanar graphs.

Proposition 8.6. *For the class of outerplanar graphs the constant y is the unique positive solution to $zB''(z) = 1$, where $B'(z) = (z + D(z))/2$ and D is given in (8.5). Moreover,*

$$\kappa = \frac{y}{2} + \left(1 - \frac{y}{2}\right) \frac{8w^4 - 16w^3 + 4w - 1}{(4w^3 - 6w^2 - 2w + 1)(2w - 1)} \approx 5.0841 \text{ with } w = D(y)$$

and $\sigma^2 = 1 + B'''(y)y^2 \approx 95.3658$.

Following [8] we develop a specification of \mathcal{B}'^{\bullet} that eventually will enable us to prove the above expressions of the relevant constants. Any 2-connected outerplanar graph has a unique Hamilton cycle, which corresponds to the boundary of the outer face in any drawing having the property that all vertices lie on the outer face. The edge set of a 2-connected outerplanar graph can thus be partitioned in two parts: the edges of the Hamilton cycle, and all other edges, which we refer to as the set of *chords*. Let \mathcal{D} denote the class obtained from \mathcal{B}' by

orienting the Hamilton cycle of each object of size at least three in one of the two directions and marking the oriented edge whose tail is the $*$ -vertex. The block consisting of a single edge is oriented in the unique way such that the $*$ -vertex is the tail of the marked edge. We start with some observations.

Lemma 8.7. *We have that $B'(z) = (z + D(z))/2$ and*

$$\mathbb{E}[\text{sp}(\Gamma B'^{\bullet}(x))] = \frac{x}{2B'^{\bullet}(x)} + (1 - \frac{x}{2B'^{\bullet}(x)})\mathbb{E}[\text{sp}(\Gamma D^{\bullet}(x))].$$

Proof. We have an isomorphism $\mathcal{B}' + \mathcal{B}' =: 2\mathcal{B}' \simeq \mathcal{X} + \mathcal{D}$. Consequently, the classes \mathcal{B}'^{\bullet} and \mathcal{D}^{\bullet} obtained by additionally rooting at a non- $*$ -vertex satisfy

$$2\mathcal{B}'^{\bullet} \simeq \mathcal{X} + \mathcal{D}^{\bullet}.$$

Hence the following procedure is a Boltzmann sampler for the class \mathcal{B}'^{\bullet} with parameter x .

$\Gamma B'^{\bullet}(x)$: $s \leftarrow \text{Bern}(\frac{x}{2B'^{\bullet}(x)})$
if $s = 1$ **then return** a single edge $\{*, 1\}$ rooted at 1
else return $\Gamma D^{\bullet}(x)$ without the orientation

This concludes the proof. \square

Hence it suffices to study the class \mathcal{D}^{\bullet} , see also Figures 5 and 6.

Lemma 8.8. *The classes \mathcal{D} and \mathcal{D}^{\bullet} satisfy*

$$\mathcal{D} = \mathcal{X} + \mathcal{D} \times \mathcal{D} + \mathcal{D} \times \mathcal{D} \times \mathcal{D} + \dots$$

$$\mathcal{D}^{\bullet} = \mathcal{X} + (\mathcal{D}^{\bullet} \times \mathcal{D} + \mathcal{D} \times \mathcal{D}^{\bullet}) + (\mathcal{D}^{\bullet} \times \mathcal{D} \times \mathcal{D} + \mathcal{D} \times \mathcal{D}^{\bullet} \times \mathcal{D} + \mathcal{D} \times \mathcal{D} \times \mathcal{D}^{\bullet}) + \dots$$

Their exponential generating functions are given by

$$(8.5) \quad D^{\bullet}(z) = \frac{z(D(z) - 1)^2}{2(D(z))^2 - 4D(z) + 1} \quad \text{and} \quad D(z) = \frac{1}{4}(1 + z - \sqrt{z^2 - 6z + 1}).$$

Proof. Let $B \in \mathcal{D}$ with $|B| \geq 2$ be a derived outerplanar block, rooted at an oriented edge \vec{e} of its Hamilton cycle C such that the $*$ -vertex is the tail of \vec{e} . Given a drawing of B such that C is the boundary of the outer face, the root face is defined to be the bounded face F whose border contains \vec{e} . Then B may be identified with the sequence of blocks along F , ordered in the reverse direction of the edge \vec{e} . This yields the decompositions illustrated in Figures 5 and 6. Solving the corresponding equations of generating functions yields (8.5). \square

The equation determining the $y = C^{\bullet}(\rho)$ is $1 = B'^{\bullet}(y) = \frac{1}{2}(y + D^{\bullet}(y))$. We obtain that $y \approx 0.17076$ is the unique root of the polynomial $3z^4 - 28z^3 + 70z^2 - 58z + 8$ in the interval $[0, \frac{1}{2}]$ and hence $\sigma^2 = 1 + B'''(y)y^2 \approx 95.3658$. It remains to compute κ .

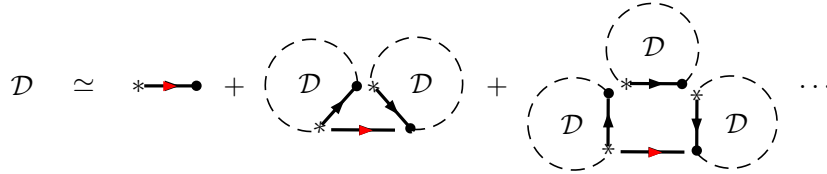


FIGURE 5. Recursive specification of the class \mathcal{D} .

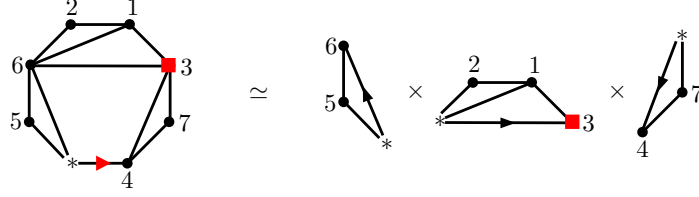


FIGURE 6. Decomposition of a \mathcal{D}^\bullet -object into a $\mathcal{D} \times \mathcal{D}^\bullet \times \mathcal{D}$ -object. The root is marked with a square.

Lemma 8.9. *We have that $\mathbb{E}[\text{sp}(\Gamma \mathcal{D}^\bullet(y))] = \frac{8w^4 - 16w^3 + 4w - 1}{(4w^3 - 6w^2 - 2w + 1)(2w - 1)} \approx 5.46545$ with $w := D(y) \approx 0.27578$.*

Since $B'^\bullet(y) = 1$ this implies with Lemma 8.7 that

$$\kappa = \mathbb{E}[\text{sp}(\Gamma \mathcal{B}'^\bullet(y))] = \frac{y}{2} + \left(1 - \frac{y}{2}\right) \mathbb{E}[\text{sp}(\Gamma \mathcal{D}^\bullet(y))] \approx 5.08418,$$

and this completes the proof of Proposition 8.6.

Proof of Lemma 8.9. The rules for Boltzmann samplers translate the specification of \mathcal{D}^\bullet given in Lemma 8.8 into the following sampling algorithm.

```

 $\Gamma \mathcal{D}^\bullet(x)$ :  $s \leftarrow$  drawn with  $\mathbb{P}(s = 2) = \frac{x}{D^\bullet(x)}$  and  $\mathbb{P}(s = i) = (i - 1)(D(x))^{i-2}$  for  $i \geq 3$ 
if  $s = 2$  then
    return a single directed edge  $(*, 1)$ 
else
     $\gamma \leftarrow$  a cycle  $\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_s, v_1\}$  with  $v_1 = *$ 
     $t \leftarrow$  a number drawn uniformly at random from the set  $[s - 1]$ 
     $\gamma \leftarrow$  identify  $(v_t, v_{t+1})$  with the root-edge of  $\gamma_t \leftarrow \Gamma \mathcal{D}^\bullet(x)$ 
    for each  $i \in [s - 1] \setminus \{t\}$ 
         $\gamma \leftarrow$  identify  $(v_i, v_{i+1})$  with the root-edge of  $\gamma_i \leftarrow \Gamma \mathcal{D}(x)$ 
    end for
    root  $\gamma$  at the directed edge  $(*, v_s)$ 
    return  $\gamma$  relabeled uniformly at random
endif

```

Given a graph H in \mathcal{D}^\bullet let $S(H)$, $S'(H)$ denote the length of a shorted path in H from the root-vertex to the tail $v_1 = *$ or head v_s of the directed root-edge, respectively. Clearly, $S(H)$ and $S'(H)$ differ by at most one. It will be convenient to also consider their minimum $M(H)$. Let S , S' and M' denote the corresponding random variables in the random graph \mathcal{D} drawn according to the sampler $\Gamma \mathcal{D}^\bullet(x)$. For any integers $\ell, k \geq 0$ with $\ell + k \geq 1$ let $\mathcal{D}_{\ell,k}$ be the random graph \mathcal{D} conditioned on the event that the graph is not a single edge and that in the root face $\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_s, v_1\}$ the length of the path $v_1 v_2 \dots v_t$ equals ℓ and the length of the path $v_{t+1} v_{t+2} \dots v_s$ equals k . Note that the probability for this event equals

$$p_{\ell,k} = \mathbb{P}(s = \ell + k + 2) \mathbb{P}(t = \ell + 1 \mid s = \ell + k + 2) = (D(x))^{k+\ell}.$$

We denote by $S_{\ell,k}$, $S'_{\ell,k}$ and $M_{\ell,k}$ the corresponding distances in the conditioned random graph $D_{\ell,k}$. Summing over all possible values for k and ℓ we obtain

$$\mathbb{E}[S] = \frac{x}{D^\bullet(x)} + \sum_{k+\ell \geq 1} \mathbb{E}[S_{\ell,k}] p_{\ell,k}, \quad \mathbb{E}[S'] = \sum_{k+\ell \geq 1} \mathbb{E}[S'_{\ell,k}] p_{\ell,k}, \quad \mathbb{E}[M] = \sum_{k+\ell \geq 1} \mathbb{E}[M_{\ell,k}] p_{\ell,k}.$$

Any shortest path from $*$ or v_s to the root-vertex of a \mathcal{B}^\bullet -graph H (\neq a single edge) must traverse the boundary of the root-face in one of the two directions until it reaches the root-edge of the attached \mathcal{B}^\bullet -object H' . From there it follows a shortest path to the root in the graph H' . Hence for all $k, \ell \geq 0$ with $k + \ell \geq 1$ the following equations hold.

$$\begin{aligned} S_{\ell,k} &\stackrel{(d)}{=} \min\{\ell + S, k + 1 + S'\}, \\ S'_{\ell,k} &\stackrel{(d)}{=} \min\{\ell + 1 + S, k + S'\}, \\ M_{\ell,k} &\stackrel{(d)}{=} \min\{\ell + S, k + S'\}. \end{aligned}$$

Since S and S' differ by at most one, this may be simplified further depending on the parameters k and ℓ as follows:

$$S_{\ell,k} \stackrel{(d)}{=} \begin{cases} \ell + S, & \ell \leq k \\ \ell + M, & \ell = k + 1, \\ k + 1 + S', & \ell \geq k + 2 \end{cases}, \quad S'_{\ell,k} \stackrel{(d)}{=} \begin{cases} k + S', & k \leq \ell \\ k + M, & k = \ell + 1, \\ \ell + 1 + S, & k \geq \ell + 2 \end{cases}$$

and

$$M_{\ell,k} \stackrel{(d)}{=} \begin{cases} \ell + S, & \ell \leq k - 1 \\ \ell + M, & \ell = k \\ k + S', & \ell \geq k + 1 \end{cases}.$$

Using this and (8.5), we arrive at the system of linear equations with parameter $w = D(x)$ and variables $\mu_S = \mathbb{E}[S]$, $\mu_{S'} = \mathbb{E}[S']$ and $\mu_M = \mathbb{E}[M]$

$$\begin{aligned} \mu_S &= \frac{2w^2 - 4w + 1}{(w - 1)^2} + \sum_{k \geq 1} \sum_{\ell=0}^k (\ell + \mu_S) w^{\ell+k} + \sum_{\ell \geq 1} (\ell + \mu_M) w^{2\ell-1} + \sum_{k \geq 0} \sum_{\ell \geq k+2} (k + 1 + \mu_{S'}) w^{\ell+k}, \\ \mu_{S'} &= \sum_{\ell \geq 1} \sum_{k=0}^{\ell} (k + \mu_{S'}) w^{\ell+k} + \sum_{k \geq 1} (k + \mu_M) w^{2k-1} + \sum_{\ell \geq 0} \sum_{k \geq \ell+2} (\ell + 1 + \mu_S) w^{\ell+k}, \\ \mu_M &= \sum_{k \geq 2} \sum_{\ell=0}^{k-1} (\ell + \mu_S) w^{\ell+k} + \sum_{\ell \geq 1} (\ell + \mu_M) w^{2\ell} + \sum_{k \geq 0} \sum_{\ell \geq k+1} (k + \mu_{S'}) w^{\ell+k}. \end{aligned}$$

Simplifying the equations yields the equivalent system $A \cdot (\mathbb{E}[S], \mathbb{E}[S'], \mathbb{E}[M])^\top = b$ with

$$A = \begin{pmatrix} 2w^4 - 4w^3 + 3w - 1 & -w^3 + w^2 & w^3 - 2w^2 + w \\ -w^3 + w^2 & 2w^4 - 4w^3 + 3w - 1 & w^3 - 2w^2 + w \\ -w^2 + w & -w^2 + w & 2w^4 - 4w^3 + w^2 + 2w - 1 \end{pmatrix}$$

and

$$b^\top = (2w^4 - 4w^3 - w^2 + 3w - 1 \quad -w \quad -w^2).$$

For $x = y \approx 0.17076$ we obtain $w \approx 0.27578$ and $\det(A) \approx -0.00799 \neq 0$. Solving the system of linear equations yields

$$\mathbb{E}[S] = \frac{8w^4 - 16w^3 + 4w - 1}{(4w^3 - 6w^2 - 2w + 1)(2w - 1)} \approx 5.46545,$$

and the proof is completed. \square

ACKNOWLEDGEMENTS

We would like to express our thanks to Grégory Miermont and Igor Kortchemski for helpful suggestions and references regarding the Continuum Random Tree. The second author would like to thank his wife for her support during the writing of this paper.

REFERENCES

- [1] L. Addario-Berry, L. Devroye, and S. Janson. Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees. *Ann. Probab.*, 41(2):1072–1087, 2013.
- [2] M. Albenque and J.-F. Marckert. Some families of increasing planar maps. *Electron. J. Probab.*, 13:no. 56, 1624–1671, 2008.
- [3] D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.
- [4] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.
- [5] D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.
- [6] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial species and tree-like structures*, volume 67 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1998. Translated from the 1994 French original by Margaret Readdy, With a foreword by Gian-Carlo Rota.
- [7] N. Bernasconi, K. Panagiotou, and A. Steger. The degree sequence of random graphs from subcritical classes. *Combin. Probab. Comput.*, 18(5):647–681, 2009.
- [8] N. Bernasconi, K. Panagiotou, and A. Steger. On properties of random dissections and triangulations. *Combinatorica*, 30(6):627–654, 2010.
- [9] J. Bettinelli. Scaling limit of random planar quadrangulations with a boundary. *To appear in Annales de l’Institut Henri Poincaré*.
- [10] P. Biane, J. Pitman, and M. Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bull. Amer. Math. Soc. (N.S.)*, 38(4):435–465 (electronic), 2001.
- [11] N. Broutin and P. Flajolet. The distribution of height and diameter in random non-plane binary trees. *Random Structures Algorithms*, 41(2):215–252, 2012.
- [12] A. Caraceni. The scaling limit of random outerplanar maps. *Submitted*, 2014.
- [13] K. L. Chung. Excursions in Brownian motion. *Ark. Mat.*, 14(2):155–177, 1976.
- [14] N. Curien, B. Haas, and I. Kortchemski. The CRT is the scaling limit of random dissections. *To appear in Random Struct. Alg.*
- [15] H. Décoste, G. Labelle, and P. Leroux. Une approche combinatoire pour l’itération de Newton-Raphson. *Adv. in Appl. Math.*, 3(4):407–416, 1982.
- [16] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, 2010.
- [17] M. Drmota, É. Fusy, M. Kang, V. Kraus, and J. Rué. Asymptotic study of subcritical graph classes. *SIAM J. Discrete Math.*, 25(4):1615–1651, 2011.
- [18] M. Drmota and M. Noy. Extremal parameters in sub-critical graph classes. In *ANALCO13—Meeting on Analytic Algorithmics and Combinatorics*, pages 1–7. SIAM, Philadelphia, PA, 2013.
- [19] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combin. Probab. Comput.*, 13(4-5):577–625, 2004.
- [20] T. Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003.
- [21] T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005.

- [22] P. Flajolet, É. Fusy, and C. Pivoteau. Boltzmann sampling of unlabelled structures. In *Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth Workshop on Analytic Algorithmics and Combinatorics*, pages 201–211. SIAM, Philadelphia, PA, 2007.
- [23] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.
- [24] O. Gimenez, D. Mitsche, and M. Noy. Maximum degree in minor-closed classes of graphs. *ArXiv e-prints*, Apr. 2013.
- [25] B. Haas and G. Miermont. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.*, 40(6):2589–2666, 2012.
- [26] F. Harary and E. M. Palmer. *Graphical enumeration*. Academic Press, New York-London, 1973.
- [27] S. Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012.
- [28] S. Janson and S. Ö. Stefánsson. Scaling limits of random planar maps with a unique large face. *To appear in Ann. Probab.*
- [29] A. Joyal. Une théorie combinatoire des séries formelles. *Adv. in Math.*, 42(1):1–82, 1981.
- [30] V. F. Kolchin. *Random mappings*. Translation Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York, 1986. Translated from the Russian, With a foreword by S. R. S. Varadhan.
- [31] I. Kortchemski. A simple proof of Duquesne’s theorem on contour processes of conditioned Galton-Watson trees. In *Séminaire de Prob. XLV*, volume 2078 of *Lecture Notes in Math.*, pages 537–558. Springer, 2013.
- [32] G. Labelle. Une nouvelle démonstration combinatoire des formules d’inversion de Lagrange. *Adv. in Math.*, 42(3):217–247, 1981.
- [33] G. Labelle. On extensions of the Newton-Raphson iterative scheme to arbitrary orders. In *22nd International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2010)*, Discrete Math. Theor. Comput. Sci. Proc., AN, pages 845–856. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010.
- [34] J.-F. Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.
- [35] J.-F. Le Gall. Itô’s excursion theory and random trees. *Stochastic Process. Appl.*, 120(5):721–749, 2010.
- [36] J.-F. Le Gall and G. Miermont. Scaling limits of random trees and planar maps. In *Probability and statistical physics in two and more dimensions*, volume 15 of *Clay Math. Proc.*, pages 155–211. Amer. Math. Soc., Providence, RI, 2012.
- [37] C. McDiarmid, A. Steger, and D. J. A. Welsh. Random graphs from planar and other addable classes. In *Topics in discrete mathematics*, volume 26 of *Algorithms Combin.*, pages 231–246. Springer, Berlin, 2006.
- [38] M. A. Méndez and J. C. Liendo. An antipode formula for the natural Hopf algebra of a set operad. *Adv. in Appl. Math.*, 53:112–140, 2014.
- [39] K. Panagiotou and A. Steger. Maximal biconnected subgraphs of random planar graphs. *ACM Trans. Algorithms*, 6(2):Art. 31, 21, 2010.
- [40] A. Rényi and G. Szekeres. On the height of trees. *J. Austral. Math. Soc.*, 7:497–507, 1967.
- [41] B. Stufler. Random enriched trees with applications to random graphs. *Manuscript in preparation.*, 2014.
- [42] G. Szekeres. Distribution of labelled trees by diameter. In *Combinatorial mathematics, X (Adelaide, 1982)*, volume 1036 of *Lecture Notes in Math.*, pages 392–397. Springer, Berlin, 1983.

(Konstantinos Panagiotou) INSTITUTE OF MATHEMATICS, LUDWIG-MAXIMILIANS-UNIVERSITÄT, THERESIENSTR. 39, 80333 MUNICH, GERMANY

E-mail address: kpanagio@math.lmu.de

(Benedikt Stufler) INSTITUTE OF MATHEMATICS, LUDWIG-MAXIMILIANS-UNIVERSITÄT, THERESIENSTR. 39, 80333 MUNICH, GERMANY

E-mail address: stufler@math.lmu.de

(Kerstin Weller) INSTITUTE OF THEORETICAL COMPUTER SCIENCE, ETH, 8092 ZURICH, SWITZERLAND

E-mail address: kerstin.weller@inf.ethz.ch